Australian Government

Department of Infrastructure, Transport, Regional Development, Communications and the Arts

BUREAU OF INFRASTRUCTURE AND TRANSPORT RESEARCH ECONOMICS
RESEARCH REPORT 156

Road

# The Economics of Road Maintenance

November 2023

BUREAU OF INFRASTRUCTURE AND TRANSPORT RESEARCH ECONOMICS

# The economics of road maintenance

Research report 156

ISBN 978-1-922879-08-0
November 2023



**An appropriate citation for this report is:**

Bureau of Infrastructure and Transport Research Economics (BITRE) 2023. *The economics of road maintenance.* Research Report 156, Canberra, ACT.

**Contact us**

This publication is available in PDF format. All other rights are reserved, including in relation to any departmental logos or trademarks which may exist. For enquiries regarding the licence and any use of this publication, please contact

Bureau of Infrastructure and Transport Research Economics
Department of Infrastructure, Transport, Cities, Regional Development, Communications and the Arts
GPO Box 594, Canberra ACT 2601, Australia

E-mail bitre@infrastructure.gov.au
Internet www.bitre.gov.au

# Foreword

Australian governments, Commonwealth, state, territory and local, spend some $30 billion per annum on the maintenance, upgrade and expansion of Australia's 875,000-kilometre road network. Maintenance expenditure is estimated to be of the order of 20 to 40 per cent of total road expenditure. This report develops analytical approaches to help ensure that road expenditure is used in the most efficient and cost-effective manner both in terms of dividing funds between construction and maintenance and allocation of maintenance funds between locations, treatment types and treatment timing.

The report focuses on the impact of timely and adequate maintenance expenditure on the overall costs to society — that is, costs to road agencies, road users and externalities. The analytical approach and case studies demonstrate the potential costs of delayed or deferred maintenance expenditure, which can result in much higher overall costs, reinforcing the adage: 'a stitch in time saves nine'. The report also provides a computer modelling approach to optimising road maintenance expenditure over time so as to minimise the overall cost to society, without and with constraints on road agency spending levels.

The report was authored by Dr Mark Harvey, who undertook the modelling and analysis. BITRE also acknowledges the input and assistance of the Australian Road Research Board (ARRB) in providing technical advice and curating the road network data set used in the case studies presented. BITRE is grateful to the Australian state government transport agency that provided the data for the case study.

Shona Rosengren

Head of Bureau
Bureau of Infrastructure and Transport Research Economics

November 2023

# At a glance

- This report discusses the economics of road maintenance, using a framework that optimises the trade-off between road agency maintenance costs and costs to road users. More frequent and more intense maintenance treatments keep a road in better condition, which reduces costs to road users, but at a higher cost to road agencies.
- For every road segment, there is a large number of possible future maintenance options involving a range of potential different treatment types, intensities and implementation times. Consequently, optimising maintenance can involve assessing a huge number of options, and is only feasible using a computer model.
- The computer model must forecast costs under each option and compare them to determine the optimum maintenance program. Models need to account for pavement deterioration, maintenance treatment costs, impacts of treatments on pavement condition, and road users' costs. Data is required on road condition, traffic volumes, maintenance treatment characteristics, environmental conditions and calibration coefficients to predict deterioration.
- BITRE developed its own model and approach to road maintenance optimisation, which includes capacity to optimise road user and agency costs within road agency budget constraints.
- A case study was undertaken using a database of 2034 road segments with a total length of 1977 kilometres drawn from the non-urban parts of seven different highways, supplied by an Australian state government road agency.
- Road agency maintenance spending is usually constrained by annual government funding allocations. The theoretical approach and BITRE model incorporate budget constraints expressed either as present values of road agency spending or as annual spending levels. Annual budget constraints act to smooth out forecast spending needs over time.
- The theoretical approach posits marginal benefit–cost ratios (MBCRs) that show the social value of increasing spending when it is constrained. As budget constraints are tightened, MBCRs rise slowly at first, then rapidly escalate. In the case of annual budget constraints, the model is able to estimate an MBCR for each individual year.
- Tight annual budget constraints in the early years of the analysis period can cause the present value of road agency costs to be higher than without the constraints due to the need for higher catch-up spending in later years. This reflects the principle of 'a stitch in time saves nine'.
- Significantly, MBCRs for maintenance, estimated via the modelling process in the report, are comparable with BCRs for capital investment projects. Such comparisons can inform decisions about the value of shifting funds between capital and maintenance budgets.
- The report also considers ways to measure the extent of underfunding of maintenance. Unless a road network is in very good condition at the start of the analysis period, an optimisation model is likely to identify a substantial 'backlog' of maintenance works that it recommends should be undertaken in the first year. The size the backlog is not a good measure of the maintenance deficit because a significant part of it is not urgent. Maintenance deficits are better measured by comparing actual current or forecast spending with a 'sustainable' or average annual forecast level of spending for a number of years into the future estimated from a model.
- Last, the report introduces the concept of 'equivalent interest rate for deferred maintenance' (EIRDM). Saving on maintenance funding in the short term in exchange for spending more in later years to repair the damage done is like borrowing money that has to be repaid later. Working out the implicit interest rate for the loan shows it can be very expensive way to borrow, with interest rates of 20% to 30% in the examples considered. The EIRDM is a way to convey to decision makers the costs of large-scale deferral of maintenance spending.

# Table of contents

# List of tables

# List of figures

# List of abbreviations

| | |
|---|---|
| **AADT** | average annual daily traffic |
| **AM** | asphalt mix |
| **ARRB** | Australian Road Research Board |
| **ATAP** | Australian Transport Assessment and Planning |
| **BCR** | benefit–cost ratio |
| **CBA** | cost–benefit analysis |
| **BITRE** | Bureau of Infrastructure and Transport Research Economics |
| **EIRDM** | equivalent interest rate for deferred maintenance |
| **GA** | genetic algorithm |
| **HDM** | highway development and management |
| **IA** | Infrastructure Australia |
| **IBCR** | incremental benefit–cost ratio |
| **IRI** | international roughness index |
| **IRR** | internal rate of return |
| **MBCR** | marginal benefit–cost ratio |
| **MESA** | millions of equivalent standard axles |
| **MGR** | maintenance gap ratio |
| **PCI** | pavement condition index |
| **PSR** | pavement serviceability rating |
| **PVAC** | present value of agency costs |
| **PVTTC** | present value of total transport costs |
| **PVUC** | present value of user costs |
| **Rehab.** | rehabilitation |
| **Resurf.** | resurface |
| **RSC** | resurface with shape correction |
| **SNP** | adjusted structural number |
| **ST** | sprayed treatment |
| **TPM** | transition probability matrix |
| **TTC** | total transport costs |
| **WTP** | willingness to pay |

# Executive summary

## Aims and scope of report (Chapter 1)

Australian governments, Commonwealth, state, territory and local, spend approximately $30 billion per annum on the maintenance, upgrade and expansion of Australia's 875,000-kilometre road network. Maintenance expenditure is estimated to be of the order of 20% to 40% of total road expenditure. There appears to be a worldwide tendency to underfund road maintenance relative to construction, attributable to financial pressures on governments combined with lower visibility to the public of maintenance spending compared to construction projects. Maintenance works are much smaller in size than construction and underspending only becomes noticeable to road users when pavement condition reaches an advanced state of disrepair (high roughness and rutting, potholes), by which time restoration costs have risen dramatically.

The report

- reviews the economics of road maintenance
- develops an approach for assessing maintenance needs at a strategic level
- tests the methodology with a case study
- suggests how the methodology could be applied to the national road network in Australia, and
- contributes to understanding the relative merits of maintenance and capital spending.

Maintenance can be defined as "all the technical and associated administrative functions intended to retain an item or system in, or restore it to, a state in which it can perform its required function" (Dekker 1996, p. 230). In common with most of the literature on road maintenance optimisation, this report focuses on periodic maintenance, which covers larger tasks undertaken at intervals of several years or more. Routine maintenance, that is, small tasks undertaken frequently, are usually costed using simple costs per lane-kilometre of road or per square metre of pavement. Only maintenance of flexible pavements (that is, pavements comprising layers of crushed rock with a waterproof seal) is considered in the report but the same broad principles apply to concrete pavements, bridges and other structures.

Decisions about periodic road maintenance involve choosing between alternative treatment types that can be applied with different intensities and at different times. If funds are scarce, a decision to treat one location can come at the expense of not treating another location. Road maintenance optimisation models seek to support decision making about maintenance by recommending a maintenance plan that will minimise or maximise an objective function usually subject to budget, resource and technical constraints. With its economic focus, the present report, concentrates on minimising the present value of costs to society as the objective function.

## Elements of road maintenance optimisation models (Chapter 2)

In road maintenance optimisation models, costs to road users are typically assumed to be a function of pavement roughness. Other important dimensions of pavement condition include cracking, pavement strength, rut depth, potholing and skid resistance. The primary drivers of increasing pavement roughness are the passing of time, climate, pavement strength and axle loads. Cracking has a significant effect on the rate of increase of pavement roughness because cracks in the bitumen seal allow moisture to penetrate the surface causing loss of pavement strength and faster deterioration.

Cracking can be prevented with resurfacing treatments applied when the bitumen oxidises and starts to become brittle. A thin resurfacing treatment, while protecting against cracking, will not reduce roughness or increase pavement strength. A thicker overlay with corrective work on pavement defects will reduce roughness and add strength. A rehabilitation treatment, that is, replacing or reworking the surface and one or more of the upper layers of the base or applying a thick overlay, will reset pavement condition parameters to the levels of a new pavement.

Pavement roughness affects user costs by reducing driver comfort, vehicle speeds and safety, and increasing fuel consumption and wear and tear on vehicles and tyres.

# Principles of road maintenance economics (Chapter 3)

## Cost minimisation approach

The economic principles that apply to road maintenance differ in some significant ways from capital investments. Economic appraisal of capital investments is undertaken via cost–benefit analysis wherein the benefits and costs to society from a project are estimated in comparison with the situation where the project is not implemented. The situation without the project is called the 'base case'. It is often described as business-as-usual or do-minimum. Several alternative options for a project may be assessed with cost–benefit analysis undertaken for each option in order to identify the best, that is, the option with the highest present value of benefits minus costs. For road maintenance, an option assessed has to involve a series of treatments over time because the economic worth of a single maintenance treatment in isolation will be affected by the timing and types of future treatments. Since there is a very large number of possible treatment type, intensity and timing combinations to choose between, the number of potential options is huge. Furthermore, there is arbitrariness in choosing which option to make the base case. The do-minimum option with routine maintenance only will eventually lead to the road deteriorating to the point where it becomes impassable. The base case therefore needs to involve some periodic maintenance treatments. The choice of the base case will affect net benefits.

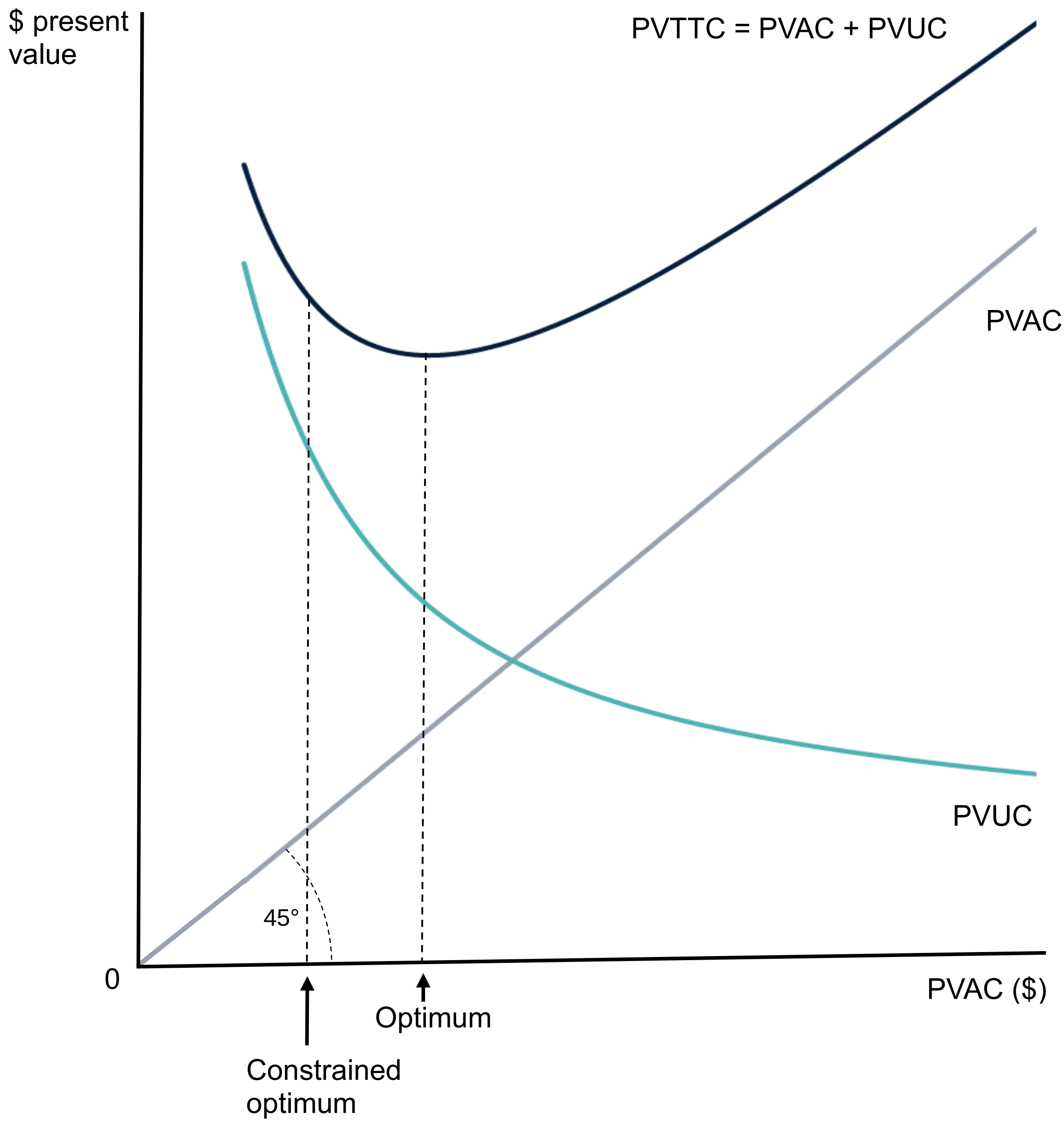
Another important difference between road maintenance and capital investment is that, while capital investment is assessed with aim of maximising benefits minus costs, maintenance can be treated as a cost minimisation problem. The reason is that, over the relevant range of road conditions, individual maintenance treatments have negligible effect on demand levels because people base their travel decision on whole-of-trip costs. Most trips will comprise travel over many road segments with pavements at different stages of their life cycles. There is therefore no need to consider benefits associated with induced traffic. It is only a matter of balancing cost savings to existing traffic from better road condition against the maintenance costs of the road agency. The optimisation problem from the point of view of society is to minimise the present value of total transport costs (PVTTC), defined as the sum of the present value of road user costs (PVUC), which may include safety and externalities, and the present value of road agency costs (PVAC).

More frequent and more intense maintenance treatments keep the road in better condition, which reduces costs to road users, but there are diminishing returns as more maintenance is undertaken. A graph could be constructed as in Figure 1 showing PVUC as function of PVAC, falling (at a decreasing rate) as PVAC is increased. PVAC as a function of itself is a 45-degree line. Adding the PVUC curve and the PVAC line together creates a U-shaped PVTTC curve. The optimum set of treatment types, intensities and timings from an economic efficiency point of view occurs where the sum of present values of costs to road users and the road agency, that is, PVTTC, is at a minimum.

## Introducing budget constraints

For capital projects in the presence of a budget constraint, project selection to maximise net benefits is achieved by ranking the projects in descending order of benefit–cost ratio (BCR). As discussed, such an approach does not transfer well to maintenance because of the huge number of options and lack of an obvious choice of base case. For maintenance, this report proposes a 'marginal benefit–cost ratio' (MBCR) defined as the saving in PVUC (benefit) from increasing PVAC (cost) by an additional dollar.

**Figure 0.1 Maintenance optimums**




At the optimum point in Figure 1, where PVTTC is at a minimum, the MBCR is one because there is no net gain from spending more. Introducing budget constraints into the optimisation model saves road agency costs at the expense of road users and raises the MBCR for additional spending (in other words, relaxing the budget constraint) above one. In Figure 1, setting a budget constraint expressed as a maximum allowable value for PVAC below the optimum, would locate a point on the PVTTC curve north-west of the optimum point. The MBCR at a point on the curve is negative the slope of the curve plus one. If the slope of the PVTTC curve at the constrained optimum point was in Figure 1 was –1.0, the MBCR would be 2.0. For large changes in maintenance spending, an incremental BCR can be calculated as the ratio of the saving in PVUC to the increase in PVAC.

## Maintenance backlog and annual budget constraints

Unless the road network is in very good condition at the start of the analysis period, with no budget constraint or with a budget constraint expressed as a present value, economically optimal maintenance spending is likely to be very high in the first year of the analysis period. Much of this 'backlog' consists of maintenance works that were economically warranted in the past but not undertaken. After the first year, economically optimal spending levels of maintenance for a network can fluctuate widely from year to year. Imposition of annual budget constraints in a model smooths spending over time, which may be needed to fit within financial and resource constraints.

Annual budget constraints on agency spending are more difficult to model than present value constraints but enable estimation of MBCRs for individual years. An annual MBCR for a given year is the present value of the user cost saving from spending an additional dollar of present value on maintenance in the given year.

MBCRs for maintenance can be an informative way to express the value of increasing maintenance spending and they have the advantage that they can be compared with BCRs for capital projects. The most economically efficient allocation of funds between maintenance and capital spending occurs where the MBCR for maintenance is the same as the cut-off BCR for capital projects.

### Cost-effectiveness analysis approach

A common alternative approach is to minimise PVAC subject to minimum road condition constraints. This is a form of cost-effectiveness analysis because it seeks to find the least-cost way to achieve the objective of maintaining roads to the specified minimum standards. Annual budget constraints can be imposed to smooth and defer road agency spending albeit at the expense of a higher PVAC value. Setting the minimum road condition objectives exogenously will lead to a less economically efficient outcome compared to PVTTC minimisation.

### Optimal pavement strength

A more general economic optimisation approach is to optimise pavement strength together with maintenance spending. It is shown that, if maintenance spending is more constrained than capital spending and funds cannot be shifted between the maintenance and capital budgets, the ‘second best’ optimum is to use some of the capital funds to build stronger pavements than would otherwise be warranted.

### Optimal incentives in maintenance contracts

Where maintenance is contracted out, an optimal maintenance outcome can be obtained with a performance-based contract in which the payment to the contractor varies negatively with road user costs.

## Maintenance optimisation modelling literature review (Chapter 4)

There is large body of literature from the civil engineering discipline on road maintenance optimisation in which authors specify a problem and present one or more solution methodologies, usually with a case study. Most models in the literature are either deterministic with continuous pavement condition or probabilistic with discrete pavement condition, adopting a Markov chain approach.

The number of possible solutions to road maintenance optimisation problems rises exponentially with the number of available treatment types, analysis years and road segments, which is known as the ‘curse of dimensionality’ or ‘combinatorial explosion’. Each case study in the literature has to manage the curse of dimensionality by restricting the numbers of segments analysed together, treatment types and analysis years, and by applying a suitable optimisation method.

Mathematical optimisation techniques such as linear, integer and dynamic programming are able to find optimum solutions provided the number of possible solutions is not too great. Heuristic optimisation algorithms, such as genetic algorithms, have extended the size of problems that can be accommodated. They can handle discrete problems and undertake random searches to find different local minimums. They can find *good* solutions but not necessarily the overall optimum solution. Their effectiveness declines for extremely large problems.

Two-stage approaches have been developed in which the best solution, or a number of best solutions, is found for each segment in isolation without any budget constraints in the first stage. In the second stage, a prioritisation approach is employed to choose a set of solutions, from among the first-stage solutions, that fits within annual budget constraints. However, prioritisation approaches are not guaranteed to find overall optimum solutions.

## Case study without annual budget constraints (Chapter 5)

A case study was undertaken using a database of 2034 segments of road with a total length of 1977 kilometres drawn from the non-urban parts of seven different highways supplied by an Australian state government road agency.

BITRE developed a maintenance model with a simplified World Road Association (PIARC) Highway Development and Management Model 4 (HDM-4) pavement deterioration algorithm. The optimisation approach involved full enumeration of all possible solutions subject to a minimum time interval between treatments. With three periodic maintenance treatment types, a minimum of eight years between treatments and a 40-year analysis period, there were 581,485 solution options for each segment. For the first 20 years, economically optimal spending for periodic (excluding routine) maintenance for the whole network was estimated at $1505 million, an average annual amount of $75 million. About 37% of the network by length would be rehabilitated over the 20 years. First-year optimal spending was estimated at $186 million, well above the annual average, suggesting a substantial maintenance backlog.

The case study data was then used to illustrate the cost–effectiveness approach of minimising the present value of road agency costs (PVAC) subject to minimum standard constraints in the form of maximum permitted roughness levels. The maximum roughness level constraint declined with traffic level and the number of heavy vehicles, patterning the relationship between traffic and economically optimal road standard. Compared with the economically optimal PVTTC-minimising solution, 20-year spending could be lowered by 18% by halving the amount of the rehabilitation work by road length in the first 10 years. However, road roughness was higher on average, imposing additional costs on road users. Each dollar of PVAC saved, on average, cost users $2.30 compared with the PVTTC-minimising economic optimum.

The unconstrained PVTTC-minimising result implies a marginal benefit–cost ratio (MBCR) of one. Optimising subject to budget constraints expressed as present values is not computationally difficult if a target MBCR above one is specified first and adjusted to give the desired PVAC value. Results were estimated for MBCRs ranging from 1.5 to 25. Increasing the MBCR pushes road agency maintenance spending into the future as well as reducing it in total. The upward adjustment to the MBCR to save each additional increment of PVAC is at first gentle as the model initially delays non-urgent treatments. But the rise soon becomes extremely steep as the model is forced to delay much-needed maintenance treatments.

A number of sensitivity tests were undertaken. Raising the discount rate reduced optimal agency spending deferring maintenance activities further into the future. Failing to include safety in user costs reduced recommended maintenance spending. If computer run times for the model need to be reduced, the sensitivity tests showed that it is better to retain the longer 40-year analysis period and skip some years in the later part of the period (that is, not to test options with treatments in the skipped years) than to shorten the analysis period to 30 or 20 years.

## Case study with annual budget constraints (Chapter 6)

Optimising maintenance subject to annual budget constraints requires all segments to be considered together. A four-stage optimisation approach was developed for the case study.

1. Full enumeration of options for all segments (already discussed)
2. Elimination of options that could not possibly appear in an optimal solution because there is a better option — that is, elimination of 'dominated' options

3. Selection of the option for each segment that minimises the objective function (PVTTC or PVAC) plus the road agency spending in each year having a budget-constraint times a 'penalty factor'. A higher penalty factor for a given year discourages selection of options with treatments in that year. The penalty factors are adjusted so that the budget constraints are just met.
4. Adjustment of the solution by allowing an 'industrial strength' genetic algorithm to select from the available options to minimise the objective function subject to the budget constraints.

Provided the penalty factors are the lowest possible to achieve the solution, MBCRs can be calculated for each year from the penalty factor for the particular year.

Case study results with annual budget constraints for the first 10 years and the first 20 years showed that

- the required penalty factors and hence MBCRs were highest for the first year when the demand for funds was greatest and declined thereafter
- a substantial proportion of the year-one maintenance backlog could be deferred at little cost provided there was optimal selection of the deferred treatments
- penalty factors, and hence MBCRs, needed to increase at an increasing rate as annual budget constraints were tightened
- as constraints were tightened, PVAC fell at first, and then rose as the costs in later years of repairing the pavements neglected during the early years began to predominate.

In summary, short deferrals of economically warranted maintenance works can be achieved at little cost with economically optimal selection of the treatments to defer, but costs to society can rise rapidly if deferral is extended. After a point, even agency costs rise in present value terms.

A simple triaging method for maintenance treatments using only a penalty factor for year one, was demonstrated to work satisfactorily for modest year-one budget constraints, but not for tight constraints.

## Measuring maintenance deficits

The first-year spending backlog is not a good measure of the maintenance deficit because a significant part of it is not urgent. Maintenance deficits are better measured by comparing actual current or forecast spending with a 'sustainable' or average annual forecast level of spending for a number of years into the future estimated from a model. A 'sustainable level' of spending can be defined as one where the jump in spending just following the constrained period is not large, or the jump is not so large that it cannot be caught up by continued spending at the sustainable level in subsequent years. Annual MBCRs can also be used to gauge the maintenance deficit, offering a measure directly comparable with the value of capital spending.

The 'equivalent interest rate for deferred maintenance' is proposed as a way to convey to decision makers the costs of deferring maintenance spending. Saving on maintenance funding in the short term in exchange for spending more in later years to repair the damage done is like borrowing money that has to be repaid later. Working out the implicit interest rate for the loan shows it can be very expensive, with interest rates of 20% to 30% in the examples considered.

## Report's contributions (Chapter 7)

The literature on road maintenance optimisation comes almost entirely out of the civil engineering discipline. While drawing heavily on the that literature, the present report was written from the perspective of the economics discipline.

The particular contributions of this report are

- explanation of the relevant road maintenance engineering concepts for non-experts
- comprehensive review of the road maintenance optimisation literature
- detailed discussion of road maintenance economics
- examination of theoretical and practical issues in road maintenance modelling
- development of the marginal and incremental BCR concepts to measure the economic value of road maintenance spending in a way that is comparable with BCRs for construction projects
- development of a methodology for optimising road maintenance for large databases of road segments. The methodology forecasts future spending needs, without or with budget constraints. Budget constraints can be expressed as present values or annual amounts. Marginal and incremental BCRs can be estimated using the methodology
- identification of data needs for modelling
- discussion of ways to simplify the methodology and reduce computer run times
- discussion of ways to estimate and communicate the size of road maintenance deficits.

The report should raise awareness of the importance of road maintenance, improve understanding of the economics of road maintenance and encourage and inform future optimisation modelling including collection of the necessary data.

# 1. Introduction

Maintenance stops infrastructure from falling into disrepair or restores infrastructure already in disrepair avoiding inconvenience and higher costs to users. Poorly maintained roads can affect vehicle operating costs and travel times, travel time reliability and safety for users. If deterioration goes too far, people will be reluctant to use the road with the attendant losses of the economic and social benefits the road confers. Under-maintaining can end up costing the road agency more in the long term because the costs of restoring the road can be much greater than the maintenance costs saved. According to Roth (2006, p. 4), delays in maintenance can double or triple the cost of subsequent road repair and reconstruction.

Underfunding of road maintenance appears to be a worldwide problem that has persisted for many years (Carnahan 1988, p. 307). Referring to developing and transition economies, Heggie and Vickers (1998, p. 42) stated that budget allocations for maintenance often fall below 50% of requirements. Maintenance funding in seven Asian countries was found to meet only 25% of national requirements (Donnges et al. 2007; World Road Association 2014, p. 14). Infrastructure Australia (IA 2016, p. 80) reported that "there is an infrastructure maintenance deficit in Australia" but was unable to estimate the size.

The absence of cost recovery arrangements to raise funds for road maintenance directly from users has been suggested as a cause of the maintenance backlog (Heggie 1995, p. 19; IA 2016, p.83 & 2019, p. 228). Being funded from general revenue, maintenance spending, in common with other government-funded activities, is subject to pressures to restrict government spending, borrowing, and taxation. Also, maintenance tends to be underfunded relative to new construction because normal preventative maintenance activities lack visibility to the public (Heggie 1995, p.24) and underspending in any year has only an incremental impact on asset condition (IA 2016, p 80). Road users have limited awareness of variations in pavement condition until it reaches an advanced state of disrepair, by which time restoration costs have risen dramatically.

## 1.1 Report objectives and scope

The report's objectives are to

- review the economic principles of road maintenance including the timing, form and quantity of maintenance
- identify an effective approach for assessing current and future spending gaps in road maintenance at a strategic level
- undertake a case study to develop and test the identified methodology
- suggest directions for a comprehensive assessment of maintenance requirements for the national road network, and
- contribute to understanding the relative merits of expenditure on maintenance of existing infrastructure and investment in new infrastructure.

Mathematical optimisation modelling plays a major role in estimating future maintenance needs. The maintenance optimisation problem is, in essence, to find the optimum balance between the costs and benefits of maintenance, while considering various constraints (Dekker 1996, p. 231). For each road segment in a network, choices have to be made between alternative treatment types and the times to implement those treatments. Where maintenance funds are limited, there is an additional problem of balancing the competing needs of the different pieces of road.

This report addresses only sealed roads with flexible pavements. They carry most vehicle-kilometres and account for most maintenance expenditure in Australia. Sealed roads with flexible pavements consist of layers of crushed rock with either a chip seal (a thin layer of bitumen and aggregate), which keeps out water, or an asphaltic concrete seal (aggregate mixed with a bitumen binder), which both keeps out water and adds structural strength. The term 'flexible pavement' refers to the fact that the pavements can deform when loads are applied and then return to their original shape.

Rigid or concrete pavements are relatively rare and relatively new, while gravel pavements are only economically warranted for low-trafficked roads. The report also does not address maintenance of bridges, tunnels, geotechnical structures, and roadside equipment. However, there are similarities between maintenance principles for different types of road infrastructure. For example, Liu et al. (1997) and Morcous and Lounis (2005) applied the same maintenance optimisation techniques to bridges and Grivas et al. (1993) to concrete pavements, that other authors apply to flexible pavements.

In contrast to the modelling for a single road pavement type that features in this report, Yeo et al. (2013, p. 318) referred to maintenance models that apply to 'heterogeneous systems' in which the 'facilities' can be of different types with different materials, deterioration processes and environmental factors. Examples are models that consider together both road segments with flexible and rigid pavements or both pavements and bridges.

## 1.2 The nature of road maintenance

Maintenance can be defined as "all the technical and associated administrative functions intended to retain an item or system in, or restore it to, a state in which it can perform its required function" (Dekker 1996, p. 230). It does not upgrade the asset. In practice, it is common to carry out minor upgrades of roads such as widening or shoulder sealing together with rehabilitations.

Road maintenance can be categorised as

- *Routine*: small tasks undertaken frequently — vegetation control, repairing or replacing signs and other roadside furniture, clearing drains and culverts, repainting line markings, patching, crack sealing and pothole repair
- *Periodic*: larger tasks undertaken at intervals of several years or more — resealing, resurfacing, overlay, reconstruction, and
- *Urgent*: unforeseen repairs requiring immediate attention — collapsed culverts, washaways, landslides that block roads (Burningham and Stankevich 2005, p. 2).

Optimisation models focus on periodic maintenance and sometimes also on components of routine maintenance that are related to roughness or the rate of pavement deterioration, in particular, crack sealing, patching and pothole repair. Road providers have considerable scope to vary the types and timing of periodic maintenance interventions. Routine maintenance, on the other hand, comprises tasks that need to be carried out if a road is to remain open to traffic and generally do not vary with traffic volume and composition. For costing purposes, routine maintenance activities that do not need to be optimised are usually assumed to be a constant amount per lane-kilometre of road or per square metre of pavement.

## 1.3 Maintenance optimisation

The focus of the present report is road maintenance *economics*, which is interpreted as decision making about maintenance activities that takes account of scarcity of resources from an economy-wide perspective. The relative scarcity of different types of resources at different times is gauged by market-determined prices or costs, which measure the value (willingness to pay) people place on the resources. Hence costs of inputs to maintenance activities such as labour, materials, fuel and equipment need to be balanced against impacts on road users' vehicle operating costs, travel time and safety. Some types of resource costs relevant to road maintenance economics have no market prices, namely, crashes and environmental externalities such as air and water pollution and greenhouse gases. However, transport project appraisal guidelines publish monetary values for these, for example, ATAP (2016).

Given resource costs, road maintenance optimisation models seek to identify the set of maintenance treatments and times at which to implement them that will achieve the best result in terms of one or more objectives within whatever constraints are imposed. Models are applied to one or more homogeneous lengths of road (segments) over a specified analysis period or planning horizon. The model has a menu of treatment types (or maintenance actions) to choose from with each treatment type having a cost and an impact on pavement condition. Input data includes the characteristics of the road segment or segments at the start of the analysis period. There has to be a sub-model that will forecast pavement deterioration in the absence of treatment. An optimisation technique is applied to find the solution.

Echoing the famous quote by the statistician George Box, "all models are wrong but some are useful," Ferreira et al. (2002a, p. 569) warned that models, by definition, do not fully capture reality. The term 'optimum' applies only to models. The optimum decisions indicated by models should help policy makers make better decisions. They support, but do not replace, the planning process and exercise of expert judgement.

One of the earliest models for maintenance optimisation, Golabi et al. (1982), was reported to have led to major cost savings. Golabi et al. (1982) developed a pavement management system for the State of Arizona to produce optimal maintenance policies for each mile of the 7,400-mile network of highways. During the first year of operation, 1980-81, the system was estimated to have saved $14 million, almost a third of Arizona's maintenance budget, and was forecast to save a further $101 million over the next four years. Two reasons for the cost reduction were identified.

> **"First, traditionally, the roads had been allowed to deteriorate to a rather poor condition before any preservation action was taken. The roads then required substantial and costly *corrective* measures. The actions recommended by the Pavement Management System are mostly *preventive* measures; that is, it recommends less substantial measures before the road deteriorates to a really poor condition. Analysis shows that less substantial but slightly more frequent measures not only keep the roads in good condition most of the time, but are overall less costly; they prevent the road from reaching really bad conditions that require much costlier corrective measures.**
>
> **Second, in the past, the corrective actions taken were too conservative; it was common to resurface a road with five inches of asphalt or concrete. The assumption was that the thicker the asphalt layer, the longer it would take for the road to deteriorate below acceptable standards. While this assumption is correct, the time it takes for a road to deteriorate is not proportional to the asphalt layer. For example, the prediction model shows that there is no significant difference between the rate of deterioration of a road resurfaced with three inches of asphalt concrete and a road resurfaced with five inches. The policies recommended by PMS [pavement management system] therefore are less conservative; for example, a recommendation of three inches of overlay is rather rare and is reserved for the worst conditions." Golabi et al. (1982, pp. 16-7)**

This suggests that the lessons learned from early efforts at road maintenance optimisation modelling have radically changed the way road agencies maintain their pavements. However, there is no confirmation from other sources. The maintenance optimisation literature is concerned with improving modelling techniques rather than assessing the practical value of maintenance modelling.

## 1.4 Project methodology

The theoretical approach in this report builds on and completes earlier BITRE research into optimising road funding including in a working paper prepared for an International Transport Forum Roundtable on Sustainable Road Funding (Harvey 2012). The working paper featured a survey of the literature on maintenance optimisation modelling and suggested ways to measure underspending on maintenance. Material from the working paper has been incorporated into the present report. Related work on road economics was published in Harvey (2015).

A case study was undertaken to test the ideas in this report and many valuable lessons were learned in the process. A database of 2034 segments of national network and state arterial roads was supplied by an Australian state government road agency. BITRE engaged the Australian Road Research Board (ARRB) to process the raw data into a form suitable for maintenance modelling, and then to estimate spending needs under several budget scenarios using the World Road Association (PIARC) Highway Development and Management Model 4 (HDM-4). BITRE then developed its own maintenance model with simplified HDM-4 pavement deterioration algorithms and ARRB's calibration factors. The flexibility and transparency of the in-house BITRE model made it possible to conduct a wide range of experiments.

The model was initially developed as an Excel spreadsheet with Visual Basic macros and subsequently recoded in Mathematica to speed up processing.

In the literature, genetic optimisation algorithms are often employed for road maintenance problems of this type. BITRE used Evolver genetic optimisation software, which links to Excel. BITRE found the genetic algorithm approach unsatisfactory when faced with an extremely large number of choices. For finding the best solution in the absence of budget constraints, BITRE used full enumeration of treatment options undertaken in Mathematica. For the more difficult subsequent task of optimising subject to annual budget constraints, a multi-stage approach was developed.

## 1.5 Structure of the report

The report is structured as follows.

- Chapter 2 introduces the technical concepts and terminology referred throughout the rest of the report, as well as the physical and economic relationships that go into maintenance modelling.
- Chapter 3 discusses the principles of road maintenance economics and introduces concepts such as the marginal and incremental benefit–cost ratios.
- Chapter 4 provides a literature review summarising the range of approaches taken by academic authors to optimise road maintenance.
- Chapter 5 presents the optimisation methodology developed for this report and the case study using actual road data for optimisation in absence of annual budget constraints.
- Chapter 6 continues the discussion and case study with the added complication of annual budget constraints.
- Chapter 7 summarises the lessons learned.

# 2. Elements of road maintenance optimisation models

## Summary

The road maintenance optimisation problem from the point of view of society involves trading off road agency or maintenance costs against road user costs over time. Three essential elements of a model designed to optimise the trade-off are

- prediction of future pavement condition (deterioration)
- prediction of the effects of specified maintenance treatments on road condition and road agency costs, and
- estimation of road user costs as a function of roughness.

*Pavement condition:* Roughness is the main dimension of pavement condition that affects road users and so is central to modelling road maintenance. Other dimensions of pavement condition featuring in models include cracking, pavement strength, rut depth, potholing and skid resistance.

In the deterioration model within the World Road Association (PIARC) Highway Development and Management Model 4 (HDM-4), based on Patterson (1987), the main drivers of roughness increase are time, climate, pavement strength and axle loads. Cracking plays a major role in roughness increase because cracks in the bitumen seal allow moisture to penetrate the surface causing loss of pavement strength and faster deterioration.

*Effects of maintenance treatments:* Cracking can be prevented with resurfacing treatments applied when the bitumen oxidises and starts to become brittle. A thin resurfacing treatment, while restoring the surface, will not reduce roughness or increase pavement strength. A thicker overlay with corrective work on pavement defects will do so. A rehabilitation treatment, that is, replacing or reworking the surface and one or more of the upper layers of the base or applying a thick overlay, will reset pavement condition parameters to the levels of a new pavement.

*Road user costs:* Pavement roughness affects user costs by reducing driver comfort and vehicle speeds, and increasing fuel consumption (and hence emissions) and wear and tear on vehicles and tyres. Roughness, rutting and skid resistance affect safety. In models requiring a simple user cost relationship, user costs are typically made a linear or quadratic function of roughness. The user cost model used in the case study later in the report makes fuel consumption, emissions, vehicle wear and tear, and safety costs a function of roughness. Speed is assumed not to be affected over the relevant range of road roughness and values of willingness-to-pay for driver comfort are not available.

## 2.1 Introduction

The topic of road maintenance economics straddles the civil engineering and economics disciplines. Application of economic principles to road maintenance requires some understanding of the technical aspects of road maintenance. This chapter explains the technical terminology and the relationships that underpin the discussion in the rest of the report. The first topic addressed is the dimensions of road condition and how they are measured. Road engineers will already be familiar with the material, but others may not. Three important components of maintenance optimisation models are the covered — prediction of how pavements will deteriorate, how different maintenance treatments improve road condition, and how road condition affects user costs. In the terminology of the HDM-4 model, these are respectively road deterioration, works effects, and road user effects. Along the way, the discussion introduces the pavement deterioration model and road user costs relationship used in the BITRE model developed to undertake the case study presented in Chapters 5 and 6 of the report.

## 2.2 Road segmentation and data

For management and modelling purposes, roads are divided into 'segments' or 'sections' that are assumed to have homogeneous characteristics. Segment length in databases can be uniform, for example, road condition measurements taken at 50 or 100 metre intervals, or variable. In the case study database, segments ranged in length from 15 metres to 18 kilometres, with an average of just under 1.0 kilometre. Data required for each segment for maintenance modelling includes

- traffic levels and growth rates for vehicles of different types
- road pavement characteristics such as pavement type, age and width
- condition data as discussed in the next section, and
- factors that affect pavement deterioration such as the environment or climate.

Although road segments are meant to be homogeneous, a certain amount of averaging of road condition over each segment is inevitable. If segments are too long with too much variation in road condition, optimal treatments based on average road condition for each segment may be inadequate for the worst parts of sections and wastefully over-treat the better parts. However, because of the set-up costs and delays to road users, short lengths of pavement in need of maintenance are unlikely to be treated without also treating less 'needy' contiguous and nearby lengths of pavement. Models rarely allow for such interdependencies.

## 2.3 Road condition measurement

The most important quantifiable attributes of road condition for strategic-level maintenance modelling are

- roughness, measured in metres per kilometre of international roughness index (IRI)
- cracking, measured as the percentage of area cracked
- rutting, measured as mean rut depth in millimetres, and
- pavement strength, measured as adjusted structural number.

### 2.3.1 Roughness

Roughness measures the ride quality of a pavement, that is, the relative comfort offered to road users. It relates to surface irregularities with wavelengths between 0.5 and 50 metres in the longitudinal profiles of either or both wheelpaths in a traffic lane. In the past, measurements were taken from the physical movement of a car's rear axle relative to its body as the vehicle travels along the road at a constant speed. The distances moved upward and downward by the axle were summed (absolute values, so downward does not offset upward) to obtain the international roughness index (IRI) in metres of vertical displacement over a kilometre travelled. Current practice is to measure the longitudinal profile of the road and to mathematically model the response of a hypothetical vehicle. Australian practice is to build into the definition a travel speed of 80 km/h and to take the average of the two wheelpaths of a lane (Austroads 2018).

### 2.3.2 Cracking

As defined in Austroads (2018, p.49), "a crack is an unplanned break or discontinuity in the integrity of the pavement surface, usually a narrow fracture or partial fracture". Cracking does not of itself add to roughness because the crack openings are narrow and easily bridged by the tyre. However, cracks allow ingress of water, which weakens pavements causing accelerated deterioration. There are different types of cracking, for example, linear (transverse of longitudinal), interconnected (crocodile or block), irregular (meandering, diagonal, crescent) and edge cracking, with varying spacing between them. The extent of cracking is measured by the percentage of surface area cracked, which can be estimated from visual inspection or automated methods involving digital cameras with image processing software.

### 2.3.3 Rutting

From Austroads (2018, p. 10), "a rut is defined as a longitudinal depression that forms in the wheelpath of a road. The length-to-width ratio would normally be greater than 4:1. Rutting may occur in one or both wheelpaths of a road." Rut depth is measured in millimetres as the maximum vertical displacement in the transverse profile. Rut depth alone does not give rise to roughness if the depth is uniform. It is the *variation* of rut depth that affects roughness. Hence, it is the standard deviation of the longitudinal profile of rut depth that appears as a contributor to road roughness in the HDM-4 deterioration model (Paterson 1987, p. 287). Rutting can be a road safety concern because, in wet weather, water 'ponds' in the ruts with potential loss of skid resistance for vehicles at high speeds. The presence of rutting can indicate inadequate pavement strength. (Austroads 2018, p.11).

### 2.3.4 Pavement strength

Pavement strength refers to the ability to carry repeated heavy axle loadings before the pavement shows unacceptable signs of structural and surface distress that seriously compromise its function (Austroads 2018, p. 21). For pavement deterioration prediction purposes, pavement strength is measured by structural number. It is a measure of the total thickness of the road pavement with each layer given a weight according to its strength, in other words, a linear combination of the layer strength coefficients and thicknesses of the individual layers above the subgrade (Morosiuk et al. 2004). The *adjusted* structural number (SNP) takes into account the contribution to pavement strength of the subgrade (the soil and rock beneath the pavement).[1] Pavement strength can be estimated from the size and shape of the depression of a pavement's surface (the 'deflection bowl') caused by a standard load, typically 40 or 50 kilo-newtons of downward force.

### 2.3.5 Other

Potholes are the most visible and severe form of pavement distress. They can cause tyre blowout, damage to wheels and suspension systems, and significantly reduce vehicle speeds and safety (Paterson 1987, p. 230). Potholes develop from wide cracks or ravelling (loss of surface material).

Skid resistance is addressed below in the section on safety impacts on road users.

Many road maintenance models use composite indexes of road condition that combine two or more condition indicators into a single measure of pavement quality such as the 'pavement condition index' (PCI) or a subjective measure such as the 'present serviceability rating' (or index) (PSR or PSI).[2]

## 2.4 Pavement deterioration modelling

Pavement age, climate, pavement strength, and axle loads all affect deterioration of pavement condition. Paterson's (1987, p. 289) incremental model of pavement deterioration is presented here to illustrate how these drivers of pavement deterioration contribute. Paterson's model is the basis of the deterioration model in HDM-4, a simplified version of which was used in the case study model in this report.

Under traditional approaches, pavement deterioration relationships can be classified as mechanistic, empirical or a combination of both. Empirical models are based on statistical (usually regression) analysis of locally-observed deterioration trends. They require extensive historical data and may not be transferrable to other locations where conditions are different (Khan et al. 2012, p. 7; Morosiuk et al. 2004, p.A2-1). The mechanistic approach relies on theory (stress, strain and deflection) to model deterioration trends. Such models use a large number of variables relating to material properties, environmental conditions, geometric elements and loading characteristics (Khan et al. 2012, p. 7). They are more easily transferable to different

----------

1 Throughout this report, the HDM-4 term 'adjusted structural number' (SNP) is used rather than the term used in earlier versions of the HDM model and Paterson (1987) of 'modified structural number' (SNC).

2 The pavement condition index (PCI) ranges from from zero (worst) to 100 (best). It is calculated as 100 minus a weighted sum of scores for different types of distress (surface defects, surface deformations and cracking) assigned by an inspector. The present serviceability rating (PSR) is a subjective scale ranging from 5 (excellent) to 0 (essentially impassable). It is the mean of the individual ratings made by a panel of experts (Carey and Irick 1960, p. 42). Through regression analysis, relationships have been developed to predict PSR from physical measurements of road condition. Since the PSR is based on passenger interpretations of ride quality, it generally reflects road roughness because roughness largely determines ride quality. According to Paterson (1986, p. 56), the relationship between PSR and the international roughness index is $PSR = 5e^{-0.18\,IRI}$.

pavements and conditions than empirical models, but are usually very data-intensive. To overcome the drawbacks of both types of model, Paterson (1987) adopted a combined mechanistic–empirical approach for the HDM-III model. This involved identifying the functional form and primary variables that affect each form of pavement deterioration from both theoretical and empirical information and then using various statistical techniques to calibrate it (Morosiuk et al. 2004, p. A2-1). Such models have moderate data requirements and can be transferred to different pavements and conditions with changed calibration parameters.

Machine learning and neural network models are now being used to predict pavement deterioration (Justo-Silva et al. 2021; Shtayat et al. 2022). These are applications of artificial intelligence. Being a recent development, it remains to be seen how they can be integrated into maintenance optimisation models.

### 2.4.1 Roughness

Paterson's (1987, p. 289) incremental model for predicting the annual change in roughness is shown in equation 2.1. 'Incremental models' predict the *change* in pavement condition over a period of time, usually a year, in contrast to 'aggregate models' that predict condition *level* at a point in time.[3] The change in roughness is given by

$$\Delta RI_t = 134\, e^{mt} SNPK^{-5.0}\, \Delta NE_4 + 0.114\, \Delta RDS + 0.0066\, \Delta CRX + 0.010\, \Delta PAT + Z_{pot} + 0.023\, RI_t\, \Delta t \qquad (2.1)$$

where

- $\Delta RI_t$ = increase in roughness over time period $t$ (m/km IRI)
- $m$ = environmental coefficient, higher in wetter areas, set at 0.023 in Paterson (1987)
- $t$ = age of pavement or overlay (years)
- $SNP$ = adjusted structural number of pavement strength
- $SNPK$ = adjusted structural number reduced for cracking = 1 + $SNP$ – 0.000758 $H$ $CRX$ where
    - $H$ is the thickness of the cracked layer (mm), and
    - $CRX$ is the area of cracking (%)
- $\Delta NE_4$ = incremental number of equivalent standard axle loads in period $\Delta t$
- $\Delta RDS$ = increase in rut depth standard deviation of both wheelpaths (mm)
- $\Delta CRX$ = increase in indexed area of cracking (%)
- $\Delta PAT$ = increase in area of surface patching (%)
- $Z_{pot}$ = dummy intercepts estimated for sections with potholing
- $RI_t$ = roughness at time $t$ (m/km IRI)
- $\Delta t$ = incremental time period of analysis (years)

In summary, the total change in roughness during a time period equals the sum of changes in roughness due to

- structural deformation (the first two terms)
    - a function of pavement strength, axle loads during the period, the environmental coefficient, and the increase in variation in rut depth during the period,
- surface defects (the third, fourth and fifth terms)

----------

[3] An example of an aggregate model is $R(t) = [R_0 + 725\,(SNP + 1)^{-4.99}\, NE_4(t)]\, e^{mt}$, which is suitable for pavements that are structurally designed for their traffic loadings and well maintained (Paterson 1987, p. 304). An incremental model has the advantage over an aggregate model that it can start from any point and so is more easily fitted to measured pavement data. Paterson and Attoh-Okine (1992, p. 104) published a modified version of the aggregate algorithm for pavements that do not have extensive distress data. The modified equation should be applied only to pavements that are maintained at low cracking levels (<30% of area). The modified equation is $R(t) = [R_0 + 263\,(SNP + 1)^{-5} NE_4(t)]\, 1.04\, e^{mt}$.

- a function of the increases during the period of cracking, patch repairs that protrude above or below the surrounding surface, and open potholes, and
- age and environmental factors (the last term)
  - a function of the environmental coefficient multiplied by the length of the time period, $\Delta t$.

Structural deformation results from plastic deformation (that is, non-reversible changes of shape in response to applied forces) in the pavement materials under the shear stresses imposed by traffic loading. It includes the effects of environmental factors on material strength and rutting behaviour under loads.

While cracking by itself does not add to roughness, the cracking term in the model, *0.0066 ΔCRX*, arises from the local or 'birdbath' depressions that often develop in a cracked area and also the effects of wide or spalled cracks, the precursor of a pothole (Paterson 1987, p. 288; Morosiuk 2004, p. B10-3).

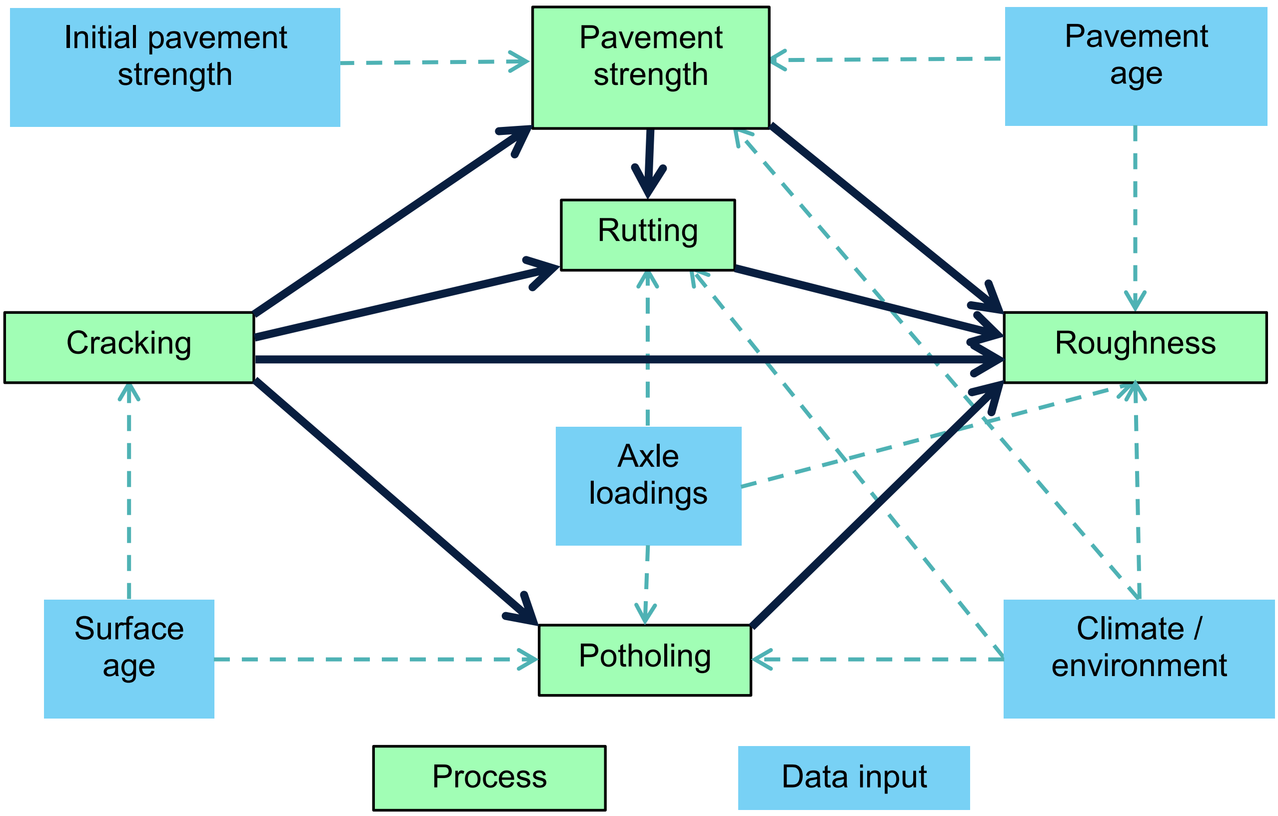
The indirect effects of cracking are serious. The bitumen surface functions as waterproofing. Cracking affects pavement strength because it allows water ingress that decreases the shear strength of granular pavement layers, which will increase deformation when the pavement is over-stressed. The rate of increase depends on material quality, the amount of water ingress (cracking and rainfall) and traffic loading (Morosiuk 2004, pp. B3-1 and B8-3). The model takes account of the effect of cracking on pavement strength by adjusting *SNP* downward to arrive at *SNPK*.

That potholes add to roughness needs no explanation. Patches add to roughness where they protrude. If the patching is performed to a high standard, the effect on roughness will be negligible (Watanatada 1987, p. 115). Effects of protruding patches are omitted altogether in the HDM-4 model.

Environmental factors, represented by the final term of Paterson's incremental model, influence roughness through non-structural effects, primarily temperature and moisture fluctuations but also foundation movements such as subsidence (Paterson 1987, p. 288).

Figure 2.1 illustrates the pavement deterioration model developed for our case study, which is a simplified version of the HDM-4 deterioration model with the addition of a relationship between pavement age and strength discussed below.

**Figure 2.1 Pavement deterioration model in case study model based on HDM-4**



Note that the figure describes the relationships within a model. It does not purport to represent the complete range of physical relationships. The roughness increase in a single year is the sum of roughness increases from four processes — cracking, pavement strength decline, potholing and rutting — plus an environmental component dependent on climate and pavement age.

The changes in pavement strength, cracking, rutting and potholing in the incremental roughness model are each determined by sub-models. Figure 2.1 also shows the main data inputs affecting each process.

## 2.4.2 Pavement strength

Strength for a newly constructed or rehabilitated pavement gradually deteriorates over time and with axle loadings. Our case study model used the relationship in equation 2.2 reported by Martin and Choummanivong (2018, p. 14) obtained from data collected at long-term pavement performance (LTPP) monitoring sites.

$$SNP_t = SNP_0\left[2 - exp\left(0.0000441333\,TMI + \frac{0.2581}{SL}t\right)\right] \tag{2.2}$$

where

- $SNP_t$ = average 'in service' adjusted structural number at age $t$
- $SNP_0$ = average 'as-built' adjusted structure number
- $t$ = age in years since construction or the last rehabilitation
- *TMI* = Thornthwaite Moisture Index, a measure of moisture deficit or surplus
- *SL* = service life, which for arterial roads is typically 60 years. (ARRB 2015, p. 20).[4]

Martin and Choummanivong were unable to obtain a relationship with both pavement age and cumulative axle loadings, probably because of the high level of multicollinearlity between them.

Pavement strength is also reduced each year by an amount related to the extent of cracking, given by *–0.000758 H CRX*, which affects *SNPK* in equation 2.1. In our model, *H* was set equal to 45.

## 2.4.3 Cracking

Age-related cracking remains at practically zero for a number of years until oxidisation of the bituminous binder reduces flexibility. In HDM-4, there is a 'crack initiation' phase for the first several years of the life of a surface during which cracking stays below 0.5% of the surface area. The length of the crack initiation phase in years depends on a calibration coefficient, traffic loading (equivalent standard axles), pavement strength and pavement type. Once the surface starts to become brittle, models assume that the percentage of the road surface cracked follows an S-curve, accelerating up to around 50%, then slowing as it approaches 100%. The shape of the S-curve varies with the calibration coefficient and pavement type.

## 2.4.4 Rutting

In the rutting sub-model of HDM-4, for a newly constructed pavement, there is a phase of 'initial densification' or 'post-construction compaction' that occurs after it is opened to traffic. Afterward, the annual increase in rutting depends on a calibration coefficient, pavement strength, equivalent standard axle loads and, after cracking commences, the level of cracking and rainfall (mean monthly precipitation).

----------

4 For estimating year-by-year progression of pavement strength decline, our case study model used the incremental form of equation 2.2, $\Delta SNP_t = exp(0.000044133\,TMI + 0.2581/SL) \cdot (SNP_t - 2SNP_0) \cdot \Delta t$. Letting $\Delta t$ equal one, the incremental equation gives the change in pavement strength over a year from the initial value of $SNP_t$ at the start of the year.

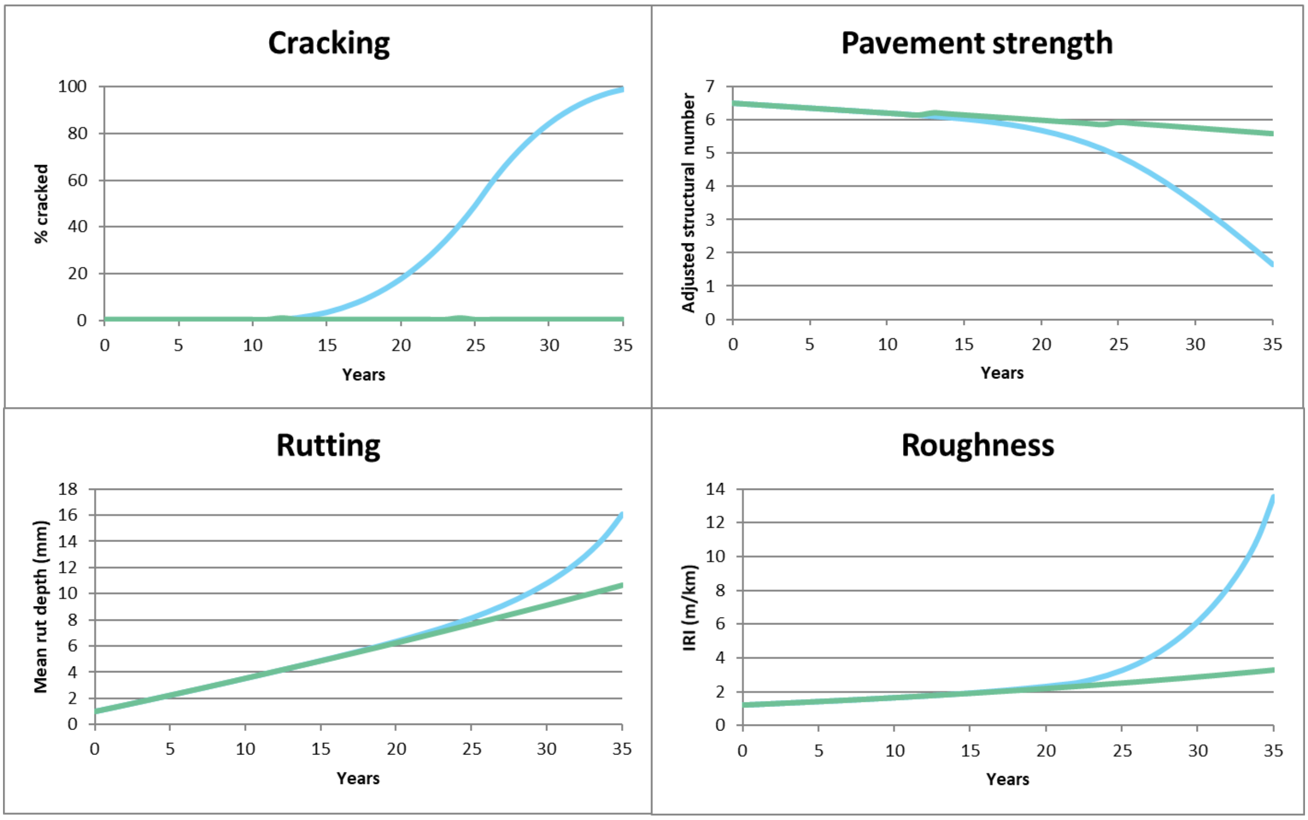
## 2.4.5 **Potholing**

In HDM-4, the annual increase in the number of pothole units (of area 0.1 $m^2$) depends on a calibration coefficient, the levels of wide cracking and ravelling, the number of axles per lane, rainfall and inversely with the thickness of the bituminous surfacing. The rate at which potholes are patched plays a major role in both pothole progression and the impact of potholes on roughness.

## 2.4.6 **Effect of not resurfacing**

With cracking impacting negatively on all the other processes in the model shown in Figure 2.1, the implication is that pavement deterioration proceeds at a higher rate following crack initiation when cracking begins to progress up the S-curve. The optimum time for a reseal or resurface is likely to be around the time of crack initiation. To illustrate the point, Figure 2.2 was created by running two scenarios in our case study model set up for a surface treatment (sprayed seal) pavement with average traffic and other characteristics in the case study database. In the scenario shown with the green lines, a resurfacing was carried out every 12 years, the length of time to crack initiation, so that cracking never rose above 0.5%. In the scenario shown with the blue lines, no resurfacings were undertaken and cracking was allowed to progress almost to 100%, a point reached after 35 years. The effects on pavement strength, rutting and roughness are clear. Note that the diagram is illustrative only. The model has been extrapolated well beyond the technical constraints normally imposed and realistic values for roughness and pavement strength.

**Figure 2.2 Pavement deterioration with (green lines) and without (blue lines) regular resurfacing**




# 2.5 Maintenance treatments

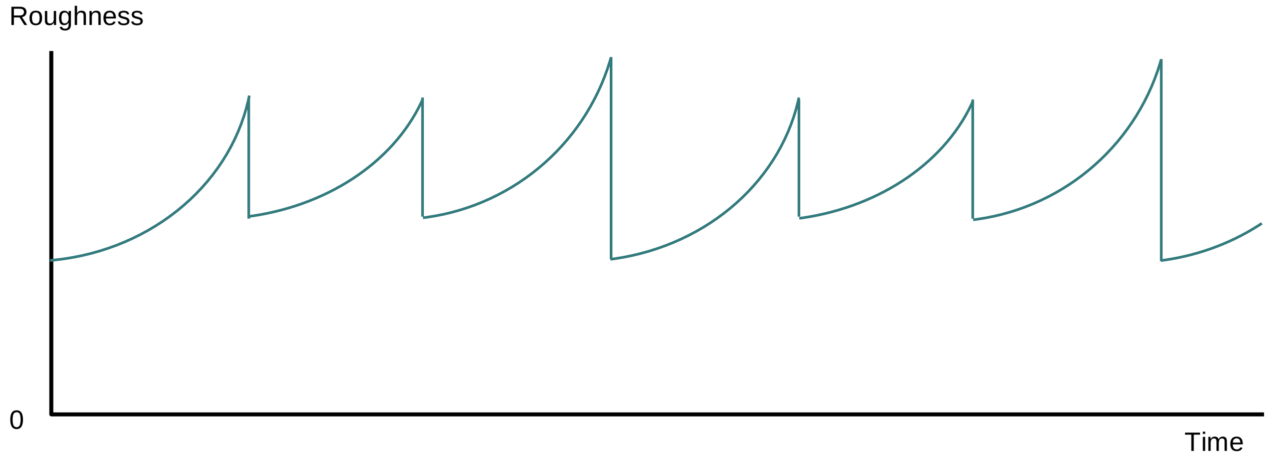
Maintenance treatments improve the levels of one or more dimensions of road condition. The characteristics of maintenance treatments will vary by location and the practices of the particular road agency. For modelling purposes, a number of treatment types may be specified. For each treatment type, the model needs to predict the effect on pavement condition. To estimate the cost of the treatment to the road agency, a cost per square metre of pavement treated is required. The cost per square metre may vary with the intensity with which the treatment is applied, for example, thicker overlays cost more, and the cost may vary with the condition of the pavement at the time of the treatment — a road in worse condition typically requires a more intense treatment.

Resurfacing treatments involve application of thin surfacings such as a layer of aggregate (gravel or crushed rock) and sprayed bitumen, or a thin layer of asphalt. They fill minor cracks, restore skid resistance, and protect the surface from aging. In a model, the time to crack initiation is reset to zero following a resurfacing treatment, but there is little or no impact on pavement strength, roughness and rutting. When combined with 'shape correction', that is, repair of surface defects and a thicker surfacing, there can be a major improvement in roughness and rutting.

Overlay or rehabilitation treatments involve replacing or reworking the surface and one or more of the upper layers of the base. Reconstruction replaces all the base layers down to the subgrade. Reconstruction effectively creates a new pavement with only the subgrade strength unchanged. (Morosiuk 2004, p. B13-13). Rehabilitation creates a new pavement as far as roughness and rutting are concerned, but the effect on pavement strength will vary with the thickness and characteristics of the overlay. Even the highest quality new asphalt surface will not have a roughness below 1 m/km IRI. Typical average roughness levels for new construction range between 1 and 2 m/km IRI and can be as high as 2.5 m/km IRI (Morosiuk 2004, p. B13-25).

Figure 2.3 illustrates the saw-tooth curve of roughness change over time in cycles of deterioration and restoration. In the diagram, rehabilitations fully restore roughness all the way to the level of a new pavement and between each rehabilitation, overlays with shape correction are applied reducing roughness part of the way to the new pavement level.

**Figure 2.3 Illustrative saw-tooth curve of pavement deterioration and restoration**




## 2.6 Road user and other societal costs

In models that optimise from the point of view of society, maintenance costs are traded-off against road user costs, that is, greater spending on maintenance can be justified by savings in user costs from pavements in better condition. Categories of user costs include, time, vehicle operation, comfort, safety and delays during maintenance.

### 2.6.1 Time and vehicle operating costs

Roughness affects road user costs in several ways. It can reduce vehicle speeds as drivers respond to decreased ride comfort and it causes wear and tear on tyres and vehicle suspension systems. Greater rolling resistance increases fuel consumption given speed.

There is no doubt that, above some critical level, roughness will reduce vehicle speeds but, due to the wide variations in research results, it is unclear what those critical levels are for cars and trucks (3 to 6 m/km IRI in the sources surveyed) nor the extent to which speed falls with roughness (McLean and Foley 1998, Opus

1999, Kalembo 2012, Parkman 2012, Yu and Lu 2014). At the upper end, Paterson and Watanatada (1985) found that travel speed is relatively insensitive to roughness at levels below 6 m/km IRI.

Where relationships have been found, the speed reduction is almost negligible over the relevant range (see the literature surveys in Parkman 2012, p.23 and Yu and Lu 2014). Yu and Lu (2014) found that average vehicle speed decreases linearly with the increase of IRI at a rate of –0.84 km/h per 1 m/km IRI. Parkman (2012, p. 178) assumed linear speed reductions of –0.59, –0.68 and –0.76 km/h for cars, light commercial vehicles and heavy commercial vehicles respectively, for each additional m/km of IRI for a study of the impacts of reduced maintenance spending in Scotland.[5]

For optimisation modelling, the relationship between roughness and road user cost is typically assumed to be either linear (for example, Li and Madanat 2002) or quadratic (for example, Ferreira and Queiroz 2012).

For the case study in the present report, the vehicle operating cost relationships for a number of vehicle types in the Australian Transport Assessment and Planning Guidelines (ATAP 2016) were used. ATAP conservatively assumes there is no relationship between roughness and vehicle speed over the relevant range of pavement roughness. The cost relationship arises from increases in fuel consumption and wear and tear on tyres and suspension systems. The relationships are quadratic with respect to roughness but the coefficient for the IRI-squared term is small compared to the coefficient for the IRI term so the relationship is close to linear. The percentage increases in vehicle operating costs (excluding time) due to roughness for the ATAP (2016) relationship over the range 1.5 to 6 m/km IRI are around 3% per IRI unit for medium-sized cars and between 6% and 8% per IRI unit for articulated trucks. By way of comparison, McLean and Foley (1998) stated that research up to that time suggested that over the range from 1.5 to 6.5 m/km IRI, road user costs excluding time rise by 4.5% for cars and 5% for articulated trucks per IRI unit.

Figure 2.4 shows the first derivatives of the user cost functions in the case study model for the five vehicle types and for safety on divided and undivided roads (discussed below). For cost minimisation, only the first derivative matters, not the absolute value of the function. In the case study model, the constants in the vehicle cost functions were set so the costs would be zero at a roughness of 1.2 m/km IRI (the assumed level for a newly rehabilitated pavement).[6,7]

### 2.6.2 **Safety**

Safety is another significant source of maintenance-related costs for road users. The relationship between safety and road condition is not often included in maintenance optimisation modelling. The literature survey in Austroads (2008) covers relationships between crash occurrence and skid resistance, microtexture, macrotexture, rutting and roughness. Pavement texture relates to wavelengths in the surface profile that are less than 50mm, much shorter than for roughness for which the wavelengths range from 0.5 to 50 metres.[8] In HDM-4, skid resistance deteriorates over time with traffic. The relationship between skid resistance and crashes is well established but studies vary in whether they consider all crashes, wet-road crashes or wet road skidding crashes. Cenek et al. (2012) found a statistically significant relationship in New Zealand data between skid resistance and crash risk.

----------

5 Based on an earlier study by Cooper (1980), Parkman (2012, p. 178) assumed that vehicles travelling on roads with IRI of 5 m/km or more have reduced vehicle speeds of 2 km/h for cars, 2.3 km/h for light goods vehicles and 2.63 km/h for heavy goods vehicles. The change in speed was made to vary linearly between these amounts at 5 m/km IRI and zero for a 3m longitudinal profile variance (LPV) of 0.5 $mm^2$. The latter converts to 1.6 m/km IRI using the conversion equation $LPV = 0.2117\ IRI^{1.8507}$ from Alonso and Yanguas (2001).

6 Similarly, Tsunokawa and Schofer (1994, p. 155), who assumed a linear user cost function with respect to roughness, did not specify a value for the constant term, noting that it was unnecessary.

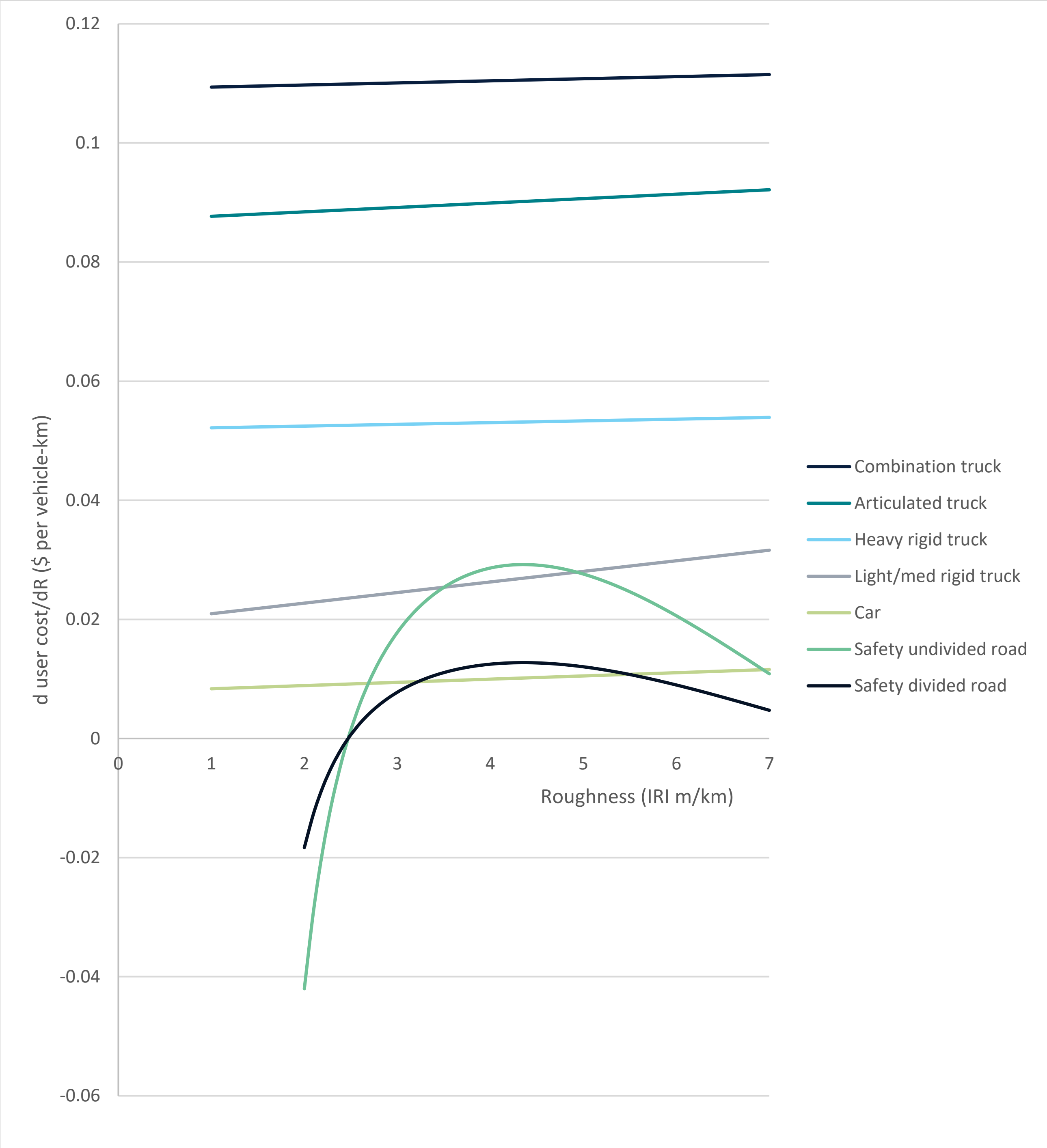
7 The coefficients for the model are: $user\ cost\ (\$/vehicle\ km) = a\ R + b\ R^2 + constant$

| | Car | Light/med rigid truck | Heavy rigid truck | Articulated truck | Combination truck |
|---|---|---|---|---|---|
| **a** | 0.007802 | 0.019169 | 0.051871 | 0.086938 | 0.109023 |
| **b** | 0.00027 | 0.000889 | 0.000146 | 0.000371 | 0.000175 |

8 Wavelength ranges for the different road surface characteristics are: micro-texture — less than 0.5mm; macro-texture — from 0.5mm to 50mm; mega-texture — from 50mm to 500mm; and roughness — from 0.5m to 50m. Austroads (2007, p. 34)

Studies of the relationship between roughness and crash risk invariably find positive correlations (Austroads 2008; Chan et al. 2009; Tehrani et al. 2017; Mamlouk et al. 2018; Lee et al. 2020). Reasons include that the contact area between tyres and the pavement decreases when pavement roughness increases, leading to lower brake friction, and diminishing the available lateral forces needed for controlling and steering vehicles (Chan et al. 2009, p.269; Tehrani et al. 2017, p.260). While slower vehicle speeds caused by roughness might be expected to reduce crash numbers and severities, drivers tend not to reduce speeds sufficiently on rougher roads.

**Figure 2.4 Rates of change in user costs as functions of roughness**



For our case study model, the relationship derived by regression analysis in Cenek et al. (2012) was used. It is a cubic function of road roughness.[9] The relationship, valid for roughness levels above 2 m/km IRI, declines slightly to reach a minimum at 2.471 m/km IRI and then increases. For our case study model, the Cenek et al. relationship was combined with crash rates from Austroads (2010) and unit costs of crashes from ATAP (2016). Presumably, the decline in crashes between 2.0 and 2.471 m/km IRI occurs because there is less skid resistance or there is a greater tendency to drive faster on smoother pavements. Since the safety relationship was not estimated for roughness values below 2.0 m/km IRI, the value was held constant at the level for 2.0 m/km IRI for lower roughness values.[10] This is consistent with Mamlouk et al. (2018), who found that crash rate is not greatly affected by roughness at low levels and estimated 'critical values' at which crashes start to increase in the range 2.4 to 4.3 m/km IRI in different U.S. states and across different time periods.

Rutting is a safety concern in wet weather when water accumulates in ruts ('ponding') increasing the risk of hydroplaning crashes. Cenek et al. (2012) found no statistically significant relationship between rut depth and crash rate and noted that they were unable to find any other studies that address the effect of rutting on crash risk. Austroads (2008), Chan et al. (2009) and Tehrani et al. (2017) similarly found little or no relationship. Mamlouk et al. (2018) was an exception but only above rut depths of 8.9 to 10mm.

Potholing can affect safety when drivers swerve to avoid potholes.

For maintenance optimisation modelling, a simpler alternative to specifying relationships between skid resistance, roughness, rut depth, potholing and crash costs is to set minimum tolerable standards that apply regardless of economic considerations. Once the standard falls below the minimum, a treatment is triggered. For example, in Toole et al. (2006), meeting minimum skid resistance standards was a major driver of surfacing treatments.

### 2.6.3 Other user costs

Other user-related impacts of road maintenance are greenhouse gas emissions, ride comfort and delays to traffic while maintenance works are carried out.

For the case study in this report, the fuel consumption model in ATAP (2016) was used to derive a relationship between roughness and the cost of additional CO2 equivalent emissions. The costs of greenhouse gas emissions added approximately one per cent to vehicle operating costs per additional IRI unit. Parkman (2012) found the impact of changes in maintenance spending on carbon dioxide emissions to be extremely small because of the offsetting additional emissions from the maintenance works and associated vehicle delays.

No values are available for users' willingness to pay for the greater comfort of travelling on smoother roads. Willingness-to-pay values to avoid discomfort due to roughness could be obtained using contingent valuation or stated preference survey methods, but it would be challenging to convey to survey respondents the feeling of driving on roads with specified roughness levels. Having them drive on lengths of road with different roughness levels would make the survey expensive to run. Moreover, correlations between roughness, vehicle operating costs, safety, and comfort would make it difficult to separately enumerate the value of discomfort.

Costs of delays to road users while maintenance activities are carried out can be significant. For local roads in Scotland, Parkman (2012, p. 64) estimated traffic delay costs to be around 25% of the level of costs of the associated maintenance works in the base case. In urban areas with high traffic levels, the need to minimise traffic delay costs affects the type of pavement laid and the times at which the works can be carried out, which adds to treatment costs. In maintenance optimisation models, these costs can be added to treatment

----------

9 The crash rate was assumed to be proportional to $Exp[-10.54 * Log_{10}R + 19.219 * (Log_{10}R)^2 - 9.85 * (Log_{10}R)^3]$ from Cenek et al. (2012, p. 30) where R is roughness in m/km IRI. In fitting the relationship to Australian crash data, it was assumed that the published casualty crash rate for undivided roads applied at a roughness of 2.5 m/km IRI and for divided roads at 2.0 m/km IRI.

10 The safety cost relationships used in the model in $ per vehicle-km were: for undivided roads $0.643134 * Exp[-10.54 * Log_{10}R + 19.219 * (Log_{10}R)^2 - 9.85 * (Log_{10}R)^3] - 0.109356$ and for divided roads $0.280396 * Exp[-10.54 * Log_{10}R + 19.219 * (Log_{10}R)^2 - 9.85 * (Log_{10}R)^3] - 0.047678$. The constants were set so the functions would be zero at the minimum value of 2.471 m/km IRI.

costs, reducing the optimal level of maintenance expenditure in the same way as would an increase in the costs of maintenance works. However, they are part of road user costs, not agency costs. The case study model undertaken for this report did not allow for costs due to maintenance works.

### 2.6.4 Additivity of user cost relationships

The user cost function for our case study was obtained by adding functions for costs of vehicle operation, CO2 emissions and safety as functions of road roughness. It would be possible to add on functions for time costs based on a roughness–vehicle speed relationship and for road user comfort as a function of roughness based on a willingness-to-pay, if satisfactory relationships were available. However, adding together relationships estimated independently can double count road user costs because of inter-relationships between different cost elements. For example, if drivers reduce speeds on rougher roads, the increased time cost will be offset by lower crash risk, wear and tear on the vehicle, and user discomfort. Caution is therefore needed when combining user cost relationships from different sources.

## 2.7 Conclusion

This chapter has introduced the technical concepts and terminology needed to understand the rest of the report. The chapter also outlined the main technical relationships that underpin the economic models of the succeeding chapters under the headings of pavement deterioration, maintenance treatments and road user costs.

Pavement deterioration models of the type discussed in this chapter require calibration to local conditions before they can be applied to actual roads. But even if well calibrated, they can at best give results that are correct on average within a wide probability distribution. The literature review in Chapter 4 shows that much road maintenance modelling is undertaken using probabilistic models. In some cases, treatment effectiveness is modelled as a probabilistic variable as well as pavement deterioration.

Road user costs are never treated as probabilistic, but the literature review just presented shows that there is a range of research findings and limited knowledge about the impacts of road roughness on users. The relationship between roughness and user costs is an area where more research is needed.

# 3. Principles of road maintenance economics


## Summary

Economic optimisation of road maintenance from the point of view of society can be treated as a cost minimisation problem rather than welfare maximisation problem. The optimisation problem is then to minimise the present value of total transport costs (PVTTC), which is the sum of the present value of road user costs (PVUC) and the present value of road agency costs (PVAC).

More frequent and more expensive maintenance treatments keep the road in better condition, which reduces costs to road users, but at a higher cost to the road agency. The decreasing cost relationship for users and the increasing cost relationship for the road agency as more maintenance is undertaken can be summed to obtain a U-shaped PVTTC curve. The set of treatment types, intensities and timings at the minimum point on the curve is the optimum maintenance policy.

A road agency budget constraint can be specified as a maximum allowable PVAC value. This leads to a constrained solution at a point on the U-shaped PVTTC curve to the north-west the minimum point.

The marginal benefit–cost ratio (MBCR) is the saving in PVUC (benefit) from increasing PVAC (cost) by an additional dollar. At the optimum point, where PVTTC is at a minimum, the slope of the curve is zero and the MBCR is one. Moving leftward along the curve as the budget constraint is tightened raises the MBCR above one.

For large changes in maintenance spending, an incremental BCR can be calculated as the ratio of the saving in PVUC to the increase in PVAC.

With no budget constraint or a present value budget constraint, optimal maintenance spending is likely to be very high in the first year of the analysis period and can fluctuate widely over subsequent years. Imposition of annual budget constraints, by smoothing spending over time, can lead to a more realistic future spending profile. An annual MBCR for a given year is the benefit from spending an additional dollar on maintenance in that year.

MBCRs for maintenance can be useful for assessing the value of increasing maintenance spending and they can be compared with BCRs for capital projects.

A cost-effectiveness analysis approach commonly applied is to minimise PVAC subject to maximum roughness constraints. Setting the maximum allowable road roughness exogenously will almost certainly lead to a less economically efficient outcome compared to PVTTC minimisation, which allows the analysis to determine the optimal maintenance standards. Present value budget constraints are not relevant but annual budget constraints can be imposed to smooth and defer road agency spending albeit at the expense of a higher PVAC value.

The theoretical discussion assumes PVTTC is minimised over an infinite period. In models, a residual value or depreciation amount can be included at the end of a finite analysis period to approximate the effect of having an infinite time horizon.

A more general optimisation approach is to optimise pavement strength and maintenance spending together, adjusting the trade-off between pavement strength and maintenance costs to minimise the sum of PVTTC for maintenance and construction costs.

Where maintenance is contracted out, an optimal maintenance outcome can be obtained with a performance-based contract in which the payment to the contractor varies negatively with road user costs. The costs incurred by users of roads with different roughness levels are thereby internalised to the supplier.

## 3.1 Introduction

The application of economics of road maintenance gives rise to a number of complexities. Periodic maintenance treatments are discrete actions that occur many years apart and the economically optimal timing for each individual treatment cannot be determined in isolation from the timing of future treatments. Then there are different treatment types and intensities of treatments to choose from, and technical and budget constraints.

For the purposes of comparing the desirability of spending on maintenance with spending on capital investments, it would be useful to be able to estimate benefit–cost ratios (BCRs) for maintenance as is done in cost–benefit analyses (CBAs) of investment projects. But with a large number of possible treatment timing and type combinations to choose from and arbitrariness in choosing the do-minimum base case, it is not straightforward to define a BCR for maintenance. It is inadequate to consider a single maintenance treatment in isolation because its value is affected by the timing and types of future treatments. Approaches are developed in this chapter for obtaining 'marginal' and 'incremental' BCRs for spending on maintenance from the optimisation process.

The chapter opens by considering the relationship between the welfare maximisation objective of CBA and the cost minimisation objective of maintenance. Then, using a simple model with a single periodic maintenance treatment type, it is shown how the optimum treatment time and maintenance standard can be determined, without and with budget constraints. Ways to define BCRs for maintenance spending are introduced in the context of the simple model. The cost-effectiveness analysis approach, minimising road agency costs subject to maximum roughness constraints is discussed, including imposition of annual budget constraints. Some complications in maintenance optimisation are then addressed, namely the effect of a finite analysis period and multiple treatment types.

Finally, two additional topics addressed using the models and concepts developed — optimising the investment-maintenance trade-off and optimal incentives in maintenance contracts. These last two topics are discussed only at a theoretical level and are not pursued further in the report.

## 3.2 Welfare maximisation versus cost minimisation

For economic analyses of road pricing and investment decisions, the optimisation problem is expressed in terms of welfare maximisation. Economic welfare derived from a road is equal to users' willingness-to-pay (WTP), the area under the demand curve for the quantity demanded, minus total social costs. Total social costs comprise road users' costs, external costs and the road agency's investment and maintenance costs. All costs should be valued at the opportunity cost of the resources consumed.[11]

For maintenance optimisation modelling, it is usual to assume the absence of a relationship between road condition and the traffic level including the vehicle mix (the proportions of the different types of cars and trucks). In effect, the demand curve for use of a road segment is assumed to be vertical over the relevant range of changes in user costs. This assumption greatly simplifies the economics of road maintenance because the benefits to additional road users do not have to be considered. The net welfare gain from an improvement in road condition is simply the resource cost saving to existing users minus the additional resource cost to the road agency and to any other members of society (external costs).

With no change in traffic, WTP becomes a constant and can therefore be omitted from the optimisation problem. The optimisation problem of maximising WTP minus the sum of road user, road agency and external costs, simplifies to minimising the sum of road user, road agency and external costs. In HDM-4 terminology,

----------

11 The term 'opportunity cost' refers to the benefit that would accrue from using a resource in its next best alternative use. 'Resource cost', also termed 'social cost' is the opportunity cost of resources used, measured from the point of view of society. Differences between private and resource costs arise when, for a given cost, the opportunities forgone are different for the individual incurring the cost and for society. Taxes, subsidies, tariffs, import quotas, unpriced externalities and non-competitive pricing by producers can cause resource costs to differ from private costs (ATAP 2022, p 16).

the sum of user, road agency and external costs is called Total Transport Costs (TTC) — terminology used throughout this report.

For short lengths of road considered in isolation, the assumption of a fixed traffic level and vehicle type mix, is usually realistic except for very high roughness levels. Road users base their demand decisions on their generalised cost for an entire trip. Most trips will comprise travel over many road segments with pavements at different stages of their life cycles. Unless an individual segment is allowed to deteriorate to the point where it can damage vehicles or significantly reduce speeds, the condition of the short individual road segment, other things held equal, should have a negligible effect on demand for road usage on the segment.

The cost minimisation approach avoids the need to specify a base case. For CBAs of investment projects, project options are always compared with a base case, usually business-as-usual or do-minimum. For maintenance, there are many alternatives to a given maintenance treatment at a given time — the same treatment can be implemented at other times and there are other treatment types and treatment intensities that can be undertaken at a range of possible times. It is possible to specify a do-minimum case against which to compare alternative scenarios of treatments and timings for same road segment. Indeed, the HDM-4 model requires it. However, there is arbitrariness in selecting a do-minimum maintenance scenario. The do-minimum option of carrying out routine maintenance only will eventually lead to the road deteriorating to the point where it becomes impassable. Some periodic maintenance treatments therefore need to be selected for the do-minimum case and there is a large range of alternatives.

Introducing a base case into the optimisation problem does not affect the optimal result. The optimisation problem: minimise the present value of total transport costs (PVTTC), or 'Minimise [*PVTTC*]', is equivalent to 'Maximise [$PVTTC_{BC} - PVTTC$]' where $PVTTC_{BC}$ is the present value of total transport costs for a do-minimum base case and is a constant.

## 3.3 Simplified example of the optimisation problem

To illustrate the principles of road maintenance economics, we use here a simplified example in which there is just one maintenance treatment type, a major rehabilitation.

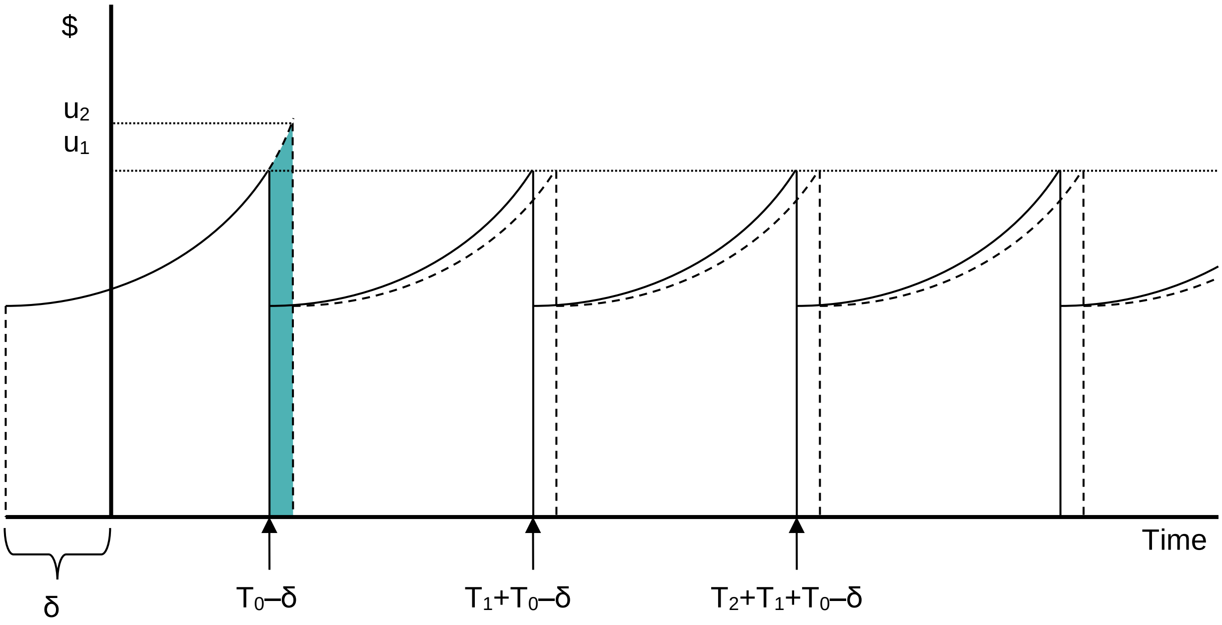
Road user costs are treated as a function of time since the last rehabilitation, $u(t)$. This function is a composite of road user costs as a function of roughness, $u = u(R)$, and roughness as a function of time, $R = R(t)$.

In Figure 3.1, the continuous lines show road user costs, $u(t)$, with rehabilitations carried out when user costs rise above a trigger level, $u_1$. The first rehabilitation is carried out when the pavement is $T_0$ years old, the second $T_1$ years later, and so on. Each rehabilitation restores roughness to its initial level in the cycle. At time zero, the pavement is $\delta$ years old, where $0 \le \delta \le T_0$. The road agency incurs rehabilitation costs in year $T_0 - \delta$, then again in year $T_1 + T_0 - \delta$, and so on.

The dashed lines in Figure 3.1 show the effect of delaying the time of the first rehabilitation by a small interval of time with no changes to the large intervals of time between the subsequent rehabilitations. During the year of the delay, immediately after time $T_0 - \delta$, users face additional costs equal to approximately $(u_1 + u_2)/2$. This additional cost to society is offset by a gain to the road agency from having rehabilitation costs in year $T_0 - \delta$ and all future years delayed by the small time interval. There are also changes in future user costs, both positive and negative in individual years. The net change to economic welfare is the present value of combined road user and road agency costs *with* the delay minus the present value of combined road user and road agency costs *without* the delay.

The optimum time to undertake the first rehabilitation can be found where the marginal cost of an additional year's delay to users equals the marginal benefit to the road agency. The same rule can be applied to determine the optimum times for all future rehabilitations.

**Figure 3.1 Effect on road user costs of delaying a rehabilitation**




To further simplify the model, it is assumed that traffic volume and the mix of vehicle types, which affect *u(t)*, are constant over time and the rehabilitation cost, *c*, is the same for all maintenance cycles. Under these assumptions, the optimum time between rehabilitations, *T*, will be the same for all cycles. In the literature, this type of model is referred to as a 'steady state' model.

The optimisation is done over an infinite time horizon. With continuous compounding, the present value of total transport cost for a cycle that commences with a rehabilitation is

$$PVTTC_{cycle} = c + \int_0^T u(t)\, e^{-rt} dt \tag{3.1}$$

With continuous compounding, the present value of a monetary amount, *a*, paid at time zero and then forever afterwards at intervals of *T* years is $\frac{a}{1-e^{-rT}}$.

The *PVTTC* over an infinite time horizon for a pavement of age *δ*, in which all cycles are identical is

$$PVTTC = \int_0^{T-\delta} u(t+\delta)\, e^{-rt} dt + e^{-r(T-\delta)} \frac{PVTTC_{cycle}}{(1-e^{-rT})} \tag{3.2}$$

The first term is the present value of user costs from year zero to year *T – δ*, the time of the first rehabilitation when the pavement reaches age *T*. The second term is the present value in year zero of rehabilitations every *T* years forever after starting in year *T – δ*, plus user costs between those rehabilitations.

The optimum cycle time is determined by

$$\frac{dPVTTC}{dT} = u(T)\, e^{-r(T-\delta)} - r\, e^{-r(T-\delta)} \frac{PVTTC_{cycle}}{(1-e^{-rT})^2} + e^{-r(T-\delta)} \frac{u(T)\, e^{-rT}}{(1-e^{-rT})} = 0$$

which reduces to

$$u(T) = r \frac{PVTTC_{cycle}}{(1-e^{-rT})} \tag{3.3}$$

The optimum occurs where

- the cost to users of extending cycle time by one year, *u(T)* (the shaded area in Figure 3.1 with the time delay set to one year), equals

- the benefit from delaying all future cycles by one year, given by the present value of total transport costs for future cycles, $\frac{PVTTC_{cycle}}{(1-e^{-rT})}$, multiplied by the discount rate. Note that multiplying a resource cost by the discount rate gives the amount the resource cost would earn if invested elsewhere for one year.

The initial pavement age, $\delta$, is irrelevant to determining the optimum cycle time. Higher rehabilitation costs, *c*, will increase $PVTTC_{cycle}$, requiring an increase in *T* to raise *u(T)* at the optimum. Thus, the more expensive it is to maintain roads, the lower will be the optimum standard of maintenance. Higher road user costs require an offsetting reduction in *T* in the optimum. Since road user costs consist of costs per vehicle times numbers of vehicles, higher traffic levels lead to higher values of *u(T)* and hence justify higher maintenance standards.

Another way to view the problem is to separate *PVTTC* into the present value of costs to users (*PVUC*) and the present value of costs to the road agency (*PVAC*) as shown in equations 3.4 and 3.5.

$$PVUC = \int_0^{T-\delta} u(t+\delta)\, e^{-rt}\, dt + \frac{e^{-r(T-\delta)}}{(1-e^{-rT})}\int_0^{T} u(t)\, e^{-rt}\, dt \qquad (3.4)$$

$$PVAC = \frac{e^{-r(T-\delta)}}{(1-e^{-rT})}c \qquad (3.5)$$

To illustrate these functions, a user cost curve as a function of time was fitted to outputs of the case study model for a one-kilometre length of road, with resurfacing undertaken at regular intervals as soon as cracking commences. Figures 3.2 and 3.3 show these curves and their sum, *PVTTC = PVUC + PVAC*, plotted against cycle time and *PVAC* respectively, with $\delta$ set to zero.

Over the relevant range of roughness values, only a small proportion of road user costs varies with roughness so a plot of total user costs against roughness would appear very flat. Since that part of user costs unaffected by roughness is irrelevant to the model, user costs were set to zero at the roughness for a new pavement of 1.2 m/km IRI.

In Figure 3.2, as cycle time (*T*) increases, *PVUC* rises and *PVAC* falls. Summing the two curves gives a U-shaped *PVTTC* curve. The optimal cycle time occurs at the minimum point on the curve, in this case, at 40 years, with *PVAC* = $335,000, *PVUC* = $498,000 and *PVTTC* = $833,000.[12]

Figure 3.3 presents the same relationships graphed against PVAC. Having PVAC on the horizontal axis instead of the time interval between treatments enables representation of maintenance options with different treatment types and time intervals between them. Cycle times below 18.2 years were omitted to avoid having to compress the scale of the vertical axis. PVAC plotted against itself is a 45-degree line. Moving to the right (spending more) implies a higher maintenance standard (shorter cycle times), the opposite of Figure 3.2.

As is usually the case with optimisation problems of this type, the U-shaped total cost curve is fairly flat in the region of the optimum. Being out by a few years on either side of the optimum imposes only a small additional cost on society. For example, if rehabilitations were undertaken at 35-year intervals, *PVTTC* would increase by $20,000 and if rehabilitations were undertaken at 25-year intervals, PVTTC would increase by $22,000.[13] However, if additional costs of this magnitude were incurred for a large number of kilometres of road, they could add up to a substantial amount.

Increasing the discount rate from 4% to 7% raises the optimum time interval between rehabilitations from 40 to 43.8 years. As a general rule, higher discount rates lead to lower optimal maintenance standards because they increase the gain from delaying maintenance spending.

----------

12 User costs in dollars per annum were given by the polynomial $2000t - 100t^2 + 2.6t^3$. The curve is a polynomial due to the safety component decreasing with roughness at low levels of roughness as discussed in Chapter 2, Section 2.6.2. The discount rate was set at 4%. Rehabilitation costs were set at $1,326,802.

13 Li and Madanat (2002, p. 533) also reported that the "objective function is rather flat near the optimal solution" and that "the discounted life time cost [*PVAC*] is not very sensitive to cycle time".

**Figure 3.2 Present values of costs graphed against time between rehabilitations**

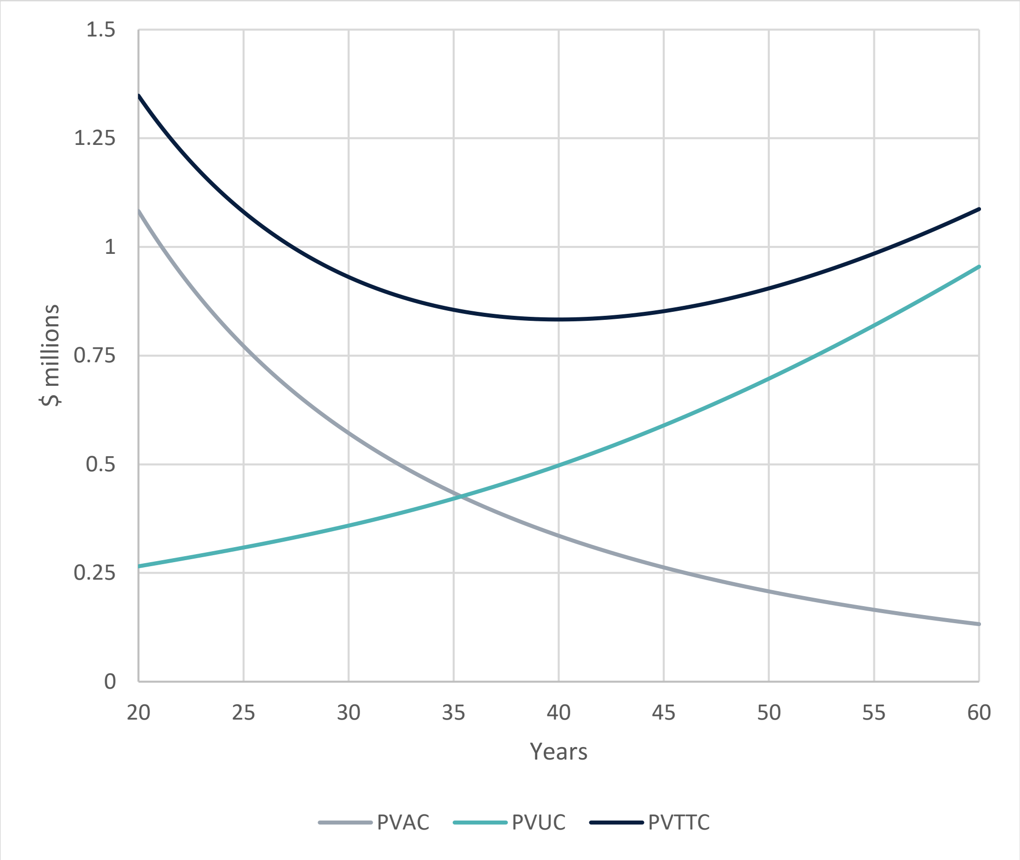
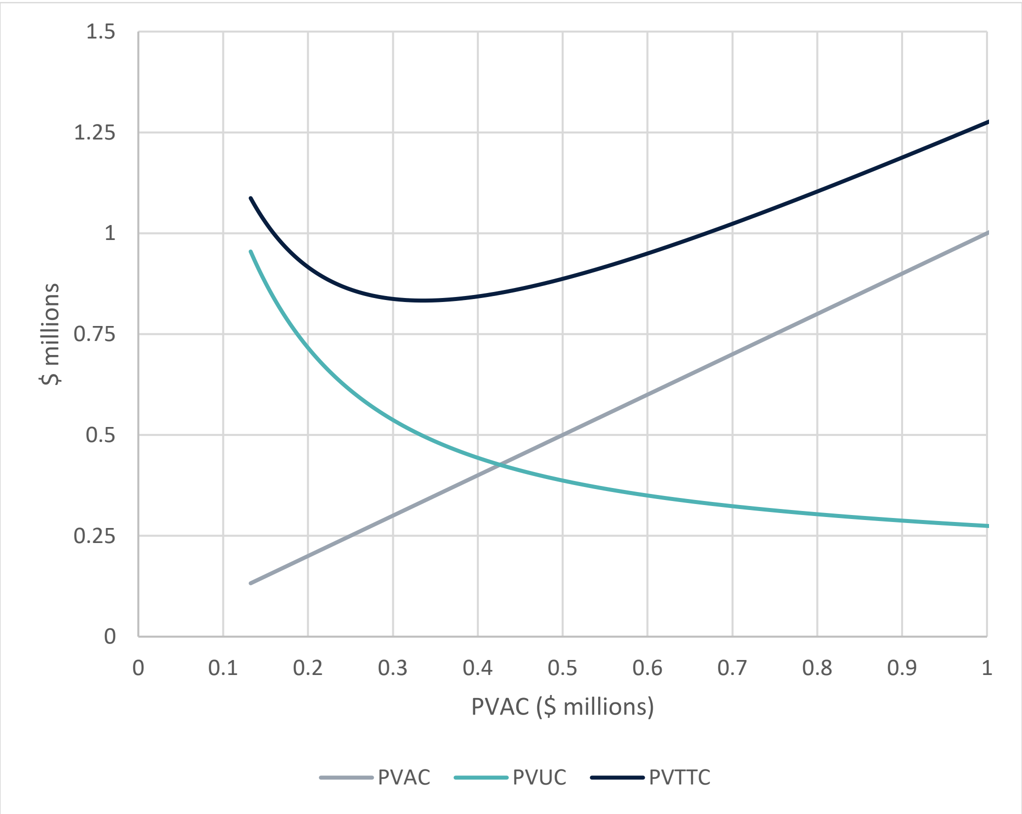



**Figure 3.3 Present values of costs graphed against agency costs**



## 3.4 The marginal benefit–cost ratio (MBCR) concept

### 3.4.1 Definition

Figure 3.3, which expresses costs as functions of PVAC instead of cycle time, suggests an alternative way to set up the optimisation problem. The aim is to find the value of PVAC that minimises PVTTC. PVUC and PVTTC can be expressed as functions of PVAC. The optimum occurs where there first derivative of PVTTC with respect to PVAC equals zero, that is

$$\frac{dPVTTC}{dPVAC} = \frac{dPVUC + dPVAC}{dPVAC} = \frac{dPVUC}{dPVAC} + 1 = 0 \qquad or \qquad -\frac{dPVUC}{dPVAC} = 1$$

The expression $-\frac{dPVUC}{dPVAC}$ is the saving in the present value of user costs (expressed as a positive number) that results from a one dollar increase in PVAC. It can be termed the marginal benefit–cost ratio (MBCR). Like a conventional BCR, it measures the benefit per dollar of additional infrastructure spending. But unlike a conventional BCR, it applies only to a very small increase in spending. The MBCR equals one at the optimum.

In our simple maintenance model with only rehabilitations occurring at intervals of *T* years, equations 3.3 and 3.4 can be differentiated and combined to express the MBCR as

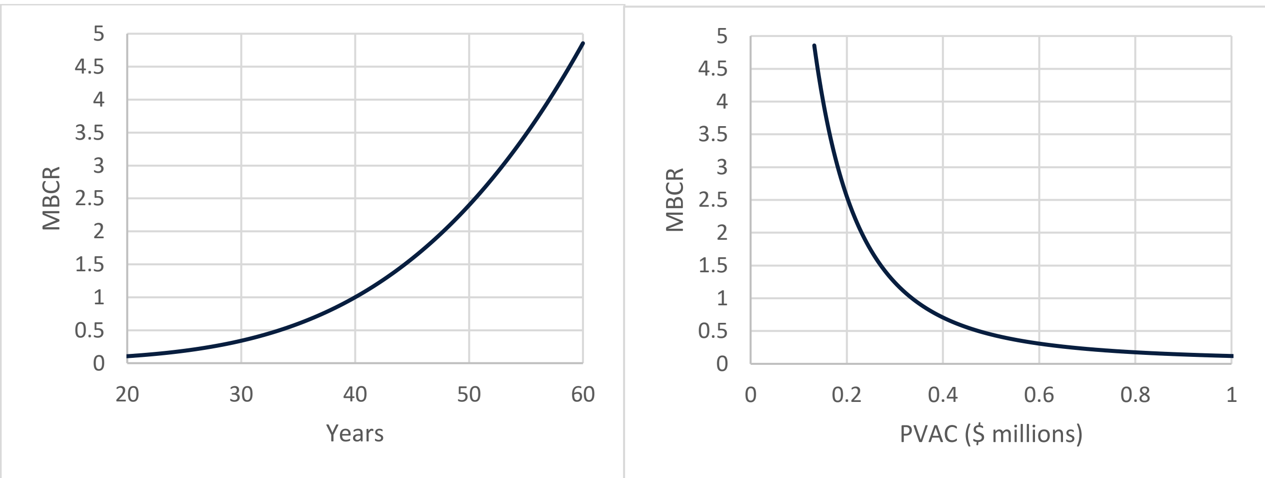
$$MBCR = -\frac{dPVUC}{dT} \Big/ \frac{dPVAC}{dT} = -\frac{dPVUC}{dPVAC} = \frac{u(T) - \dfrac{r\int_0^T u(t)\, e^{-rt}\, dt}{(1-e^{-rT})}}{\dfrac{r\,c}{(1-e^{-rT})}} \qquad (3.6)$$

Substituting in the condition for the optimum cycle time, equation 3.3, makes equation 3.6 equal one.

The two graphs in Figure 3.4 show the MBCR in our numerical example as calculated from equation 3.6 plotted against cycle time and against the present value of maintenance costs (PVAC).

The MBCR rises as cycle time increases and falls as maintenance spending (PVAC) increases. The optimum cycle time and PVAC can be read off the graphs at the points where the MBCR equals one. The MBCR is quite sensitive to non-optimal maintenance timing, being 0.60 for a 35-year cycle time and 1.59 for a 45-year cycle time. To illustrate the interpretation of the concept, the latter MBCR of 1.59 implies that, if rehabilitations were being undertaken at 45-year intervals, road users would gain (or PVUC would be reduced by) $1.59 from spending an additional dollar of PVAC to shorten cycle lengths.

**Figure 3.4 Marginal benefit–cost ratio plotted against cycle time and maintenance spending**



Equation 3.7 shows that the MBCR is negative the slope of the PVTTC curve in Figure 3.3 plus one.

$$MBCR = -\frac{dPVUC}{dPVAC} = -\frac{(dPVUC + dPVAC)}{dPVAC} + \frac{dPVAC}{dPVAC} = -\frac{dPVTTC}{dPVAC} + 1 \quad (3.7)$$

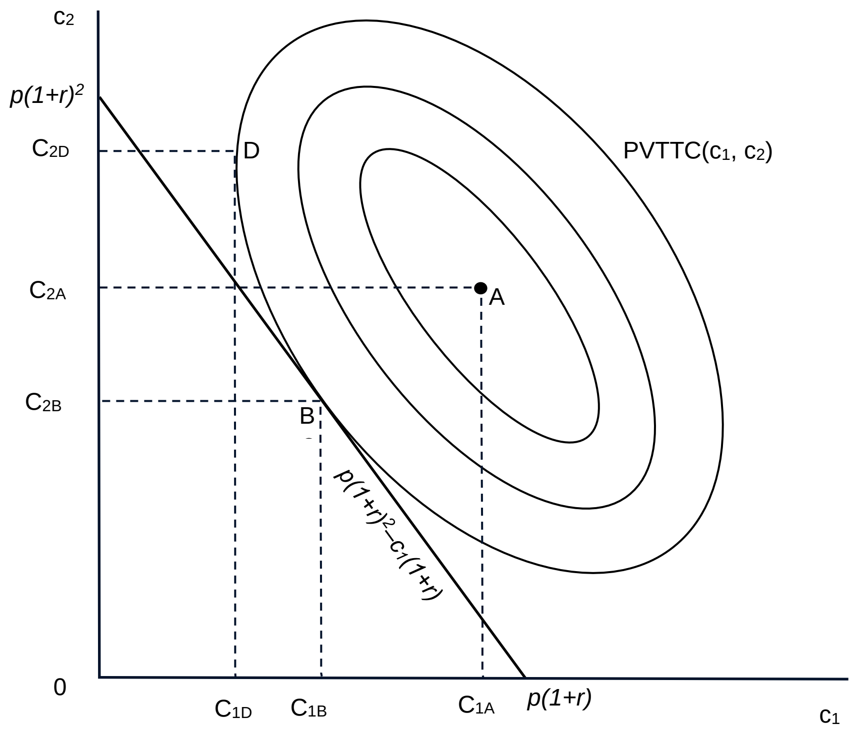
## 3.4.2 Present value budget constraints

The simplest type of budget constraint from an analytical viewpoint is a maximum allowable present value of road agency costs. The optimisation problem becomes: minimise PVTTC subject to $PVAC \leq B$, where $B$ is the maintenance budget expressed as a present value.

A present value budget constraint implies that funds can be shifted through time by borrowing or lending at an interest rate equal to the discount rate. While not necessarily realistic, it warrants discussion because it ensures an optimal allocation of limited funds over time and introduces the relationship between budget constraints and MBCRs before moving on to the more complex discussion of annual budget constraints.

Consider a single road segment in isolation for which the amount of maintenance funds spent in each year is a continuous variable. Although this assumption is unrealistic for a single segment, as demonstrated below in Chapter 6, it holds approximately for analysis of a large number of small segments taken together. Figure 3.5 shows level curves for PVTTC as a function of maintenance spending in two periods, $c_1$ and $c_2$, $PVTTC(c_1, c_2)$. In the absence of any budget constraints, optimum spending for the two periods occurs at the minimum point $A$ with spending in each of the two periods of $C_{1A}$ and $C_{2A}$ respectively.

**Figure 3.5 Unconstrained and constrained optimums**




Since PVTTC is the same for all points on a level curve, along any curve

$$dPVTTC = \frac{\partial PVTTC}{\partial c_1} dc_1 + \frac{\partial PVTTC}{\partial c_2} dc_2 = 0$$

From this, the slope of a level curve at any point is

$$\frac{dc_2}{dc_1} = -\frac{\frac{\partial PVTTC}{\partial c_1}}{\frac{\partial PVTTC}{\partial c_2}}$$

A present value budget constraint in a two-period model, PVAC = *p* implies that $p = \frac{c_1}{(1+r)} + \frac{c_2}{(1+r)^2}$, which is shown in Figure 3.5 as the line $c_2 = p(1+r)^2 - c_1(1+r)$. The constrained optimum occurs at point *B* (spending at $C_{1B}$ and $C_{2B}$) where the constraint is tangent to the lowest level curve within reach. At this point, the slope of the level curve is the same as slope of the budget constraint line at *–(1 + r)*.

A higher discount rate will make the budget constraint line steeper, moving point B north-west and shifting spending from year one to year two.

A formal mathematical approach is set out in Appendix A.1 where the optimisation problem is expressed as

Minimise $PVTTC(c_1, c_2, c_3, \ldots)$ subject to $PVAC \leq B$

where $c_t$ is maintenance spending in year *t* and *B* is the present value budget constraint.

The problem can be addressed using the method of Lagrange multipliers by minimising the Lagrange function or 'Lagrangian'

$$L = PVTTC(c_1, c_2, \ldots, c_t, \ldots) + \lambda\left[\sum_{t=1}^{\infty} \frac{c_t}{(1+r)^t} - B\right] \quad (3.8)$$

where $\lambda$ is the Lagrange multiplier, and $PVAC = \sum_{t=1}^{\infty} \frac{c_t}{(1+r)^t}$

The Lagrange method to find the maximum or minimum of an objective function $f(\boldsymbol{x})$ subject to an equality constraint $g(\boldsymbol{x}) = 0$, where $\boldsymbol{x}$ is a vector of variables, combines the two functions into a Lagrangian function, $L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$. The solution occurs at the point where all the partial derivatives of the Lagrangian function equal zero, including the derivative with respect to $\lambda$, which is the constraint, $g(\boldsymbol{x})$. The constrained optimum occurs at a saddle point of the Lagrange function. At this point, $\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} + \lambda \frac{\partial g}{\partial x_i} = 0$, or $\frac{\partial f}{\partial x_i} = -\lambda \frac{\partial g}{\partial x_i}$ for all $x_i$. This implies that the gradient vector of *f(**x**)*, the vector of partial derivatives, is proportional to the gradient vector of the constraint, that is, $\nabla f(\boldsymbol{x}) = -\lambda \cdot \nabla g(\boldsymbol{x})$, with $-\lambda$ as the proportionality constant. The gradient vectors are therefore parallel. In the two-dimensional case of Figure 3.5, at point *B*, the gradient vectors are perpendicular to the tangents of the level curves. So having proportional gradient vectors for $PVTTC(c_1, c_2)$ and $c_2 = p(1+r)^2 - c_1(1+r)$ implies that the level curves have the same slope at the optimum point, which Figure 3.5 illustrates.

There are infinitely many points for which $\nabla f(\boldsymbol{x}) = -\lambda g(\boldsymbol{x})$, or, in the case of Figure 3.5, $\frac{dc_2}{dc_1} = -(1+r)$ for differing sizes of budget constraint. Finding the desired tangent point where $g(\boldsymbol{x}) = 0$, is determined by fixing the value of $\lambda$. If there are *n* variables *x*, there are *n + 1* partial derivatives of the Lagrangian, including the partial derivative with respect to $\lambda$. Solving for the *n + 1* unknowns gives the optimum solution.

Appendix A.1 derives the mathematical conditions for the optimum for equation 3.8 and shows that, at the constrained optimum point,

$$\lambda = -\frac{dPVTTC}{dPVAC}$$

Since the Lagrange multiplier at the optimum point is the ratio of the gradient vector for the objective function to the gradient vector for the constraint, it indicates the change in the objective function that will occur for a one unit change in the constraint.

Combining this last result for $\lambda$ with equation 3.7 gives the result

$$MBCR = \lambda + 1$$

showing the connection between the Lagrange multiplier and the MBCR. Optimisation subject to a present value budget constraint can give rise to an estimate of the MBCR. The link between the Lagrange multipliers and the MBCR is discussed further below in the context of annual budget constraints.

In practical applications, finding the optimum set of maintenance treatments subject a present value budget constraint is not much more difficult than unconstrained optimisation. The Lagrangian in equation 3.8 can be written as

Minimise

$$PVTTC(c_1, c_2, \dots, c_t, \dots) + \lambda \sum_{t=1}^{\infty} \frac{c_t}{(1+r)^t}$$

$$= PVUC + PVAC + \lambda^* \times PVAC$$

$$= PVUC + PVAC + (MBCR^* - 1)PVAC$$

$$= PVUC + MBCR^* \times PVAC$$

where $\lambda^*$ and *MBCR** are target values specified by the analyst. The budget amount, *B*, can be dropped from the minimisation expression because it is a constant. The analyst needs to specify a target MBCR above one and minimise the weighted value of PVTTC, that is, $PVUC + MBCR^* \times PVAC$. The optimisation may have to be done several times to find the target MBCR value that reduces PVAC to the budget constraint.

Another, simpler way to demonstrate that minimising weighted PVTTC leads to the optimum result consistent with the target MBCR is as follows. The present values can be specified as functions of treatment time, as in equations 3.4 and 3.5. The optimisation problem is

Minimise $Weighted\ PVTTC = PVUC + MBCR^* \times PVAC$

$$\frac{dWeighted\ PVTTC}{dT} = \frac{dPVUC}{dT} + MBCR^* \frac{dPVAC}{dT} = 0$$

from which,

$$-\frac{dPVUC}{dT} \Big/ \frac{dPVAC}{dT} = -\frac{dPVUC}{dPVAC} = MBCR^*$$

Where the budget constraint applies to a group of road segments taken together or a network, the optimal allocation of maintenance funds would be found where the MBCR is the same for all segments. If one segment has a higher MBCR than another, shifting maintenance funds from the low-MBCR segment to the high-MBCR segment will generate a net saving in user costs for the two segments at no additional cost to the road agency. Hence, with a present-value budget constraint, optimisation modelling for multiple segments is not more difficult than for a single segment because one can optimise each segment in isolation from the others, applying the same target MBCR value as a weight to each segment.

In remote and regional areas, road agencies may wish to maintain low volume roads at above economically optimal standards for social reasons. To inform such a policy, a maintenance optimisation model could be run with a target MBCR below one.

### 3.4.3 Incremental BCR

The MBCR is an 'instantaneous' BCR value as it applies at single point on the PVTTC curve. It is possible to define an 'incremental' BCR (IBCR) between any two points on the curve as

$$IBCR = -\frac{\Delta PVUC}{\Delta PVAC} = -\frac{\Delta PVUC + \Delta PVAC - \Delta PVAC}{\Delta PVAC} = -\frac{\Delta PVTTC}{\Delta PVAC} + 1 = -\frac{PVTTC_1 - PVTTC_2}{PVAC_1 - PVAC_2} + 1$$

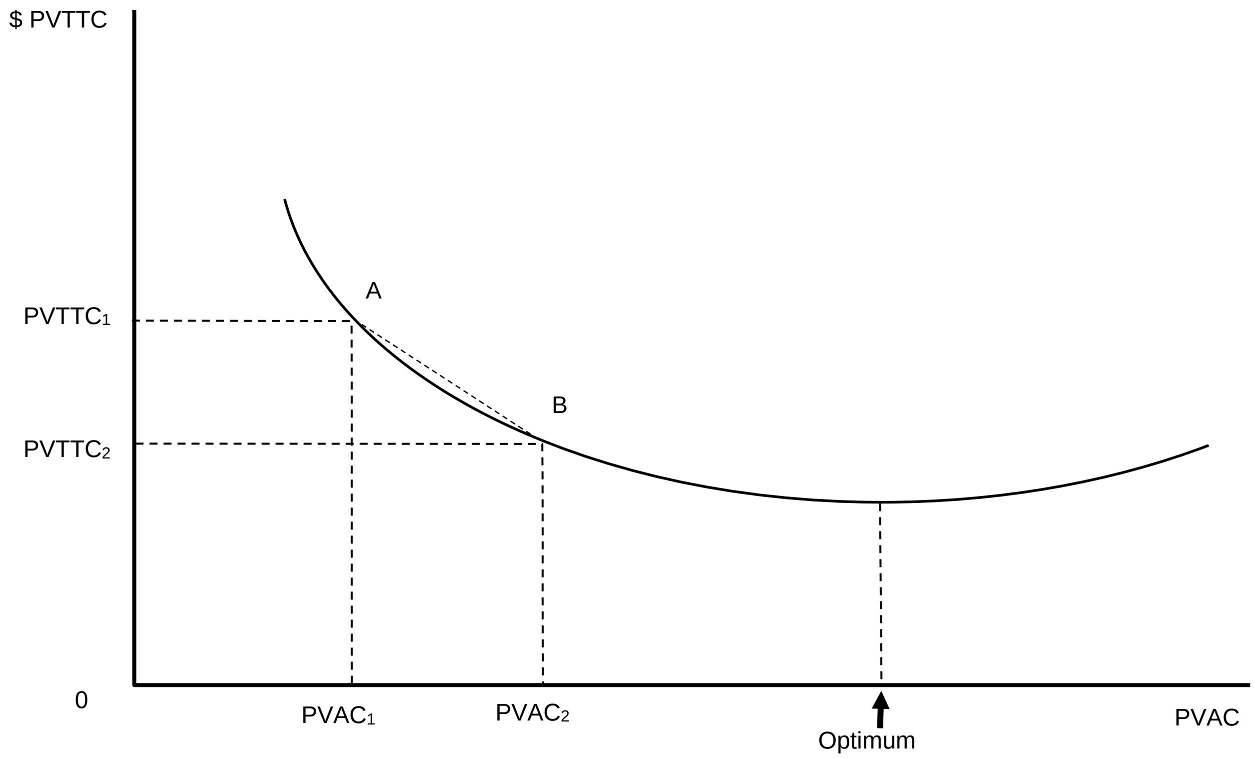
The leftmost point on the curve, subscript 1, could be a do-minimum base case.

The IBCR is one minus the slope of a line connecting the two points on the *PVTTC* curve being compared. This is illustrated in Figure 3.6 for the points A and B. Provided the curve is smooth, the IBCR between two points will lie between the MBCRs at the two points. In our numerical example, the IBCR from increasing *PVAC* from \$230,727 where the MBCR is 2.0, to \$270,891, where the MBCR is 1.5, is

$$IBCR = -\frac{\$877{,}329 - \$847{,}873}{\$230{,}727 - \$270{,}891} + 1 = -\frac{-\$29{,}456}{\$40{,}164} + 1 = 1.73$$

It is important that the PVTTC values at the start and endpoints for an IBCR be on the curve, that is, at the lowest possible PVTTC value for each PVAC value. For example, if the first point was above the curve and the second on the curve, the benefit from optimisation will be counted together with the benefit from the funding increase. The benefit from finding the optimal solution for a given PVAC value could be obtained without any change in PVAC.

**Figure 3.6 Incremental BCR**




## 3.4.4 Annual budget constraints

Many road maintenance optimisation applications involve annual budget constraints. As noted above, present value budget constraints imply that the road agency can shift funds through time by borrowing and lending at the discount rate, which may not be realistic. Furthermore, unless the network is in very good condition at the start of the analysis period, the model is likely to recommend a very large amount of spending in the first year to catch up any backlog. Such a large amount in a single year may not be financially or physically possible. The economically optimal amounts of spending may fluctuate in subsequent years as well. Indeed, Fwa et al. (1994b, p. 713) consider a long-term uniformly distributed maintenance demand to be a meaningful objective of a maintenance program.

Typically, in the modelling literature, a budget constraint for a whole network of segments together is set for each of the first several years, then no constraints thereafter. For example, Archondo-Callao (2008), demonstrating the HDM-4 model, imposed uniform budget constraints for the first five years of the analysis period.

Referring back to the two-period case illustrated in Figure 3.5, when the first-year budget $c_1$ is constrained to $C_{1D}$, the constrained optimum is found at point *D* where a vertical line from $C_{1D}$ is tangent to a level curve at point *D*. With no constraint on year-two spending, $\frac{\partial PVTTC}{\partial c_2} = 0$, and the slope of the level curve at point D is infinite. At point *D*, as well as spending in year one being below the unconstrained optimum as required to meet the constraint, spending in year two is higher than at the unconstrained optimum. Generally, the tighter the constraint on first year spending, the higher the optimal amount of spending in the unconstrained second year. Thus, clamping down on spending in a constrained year, pushes maintenance costs into unconstrained years. This effect is strongly in evidence in the case studies with annual budget constraints presented in Chapter 6.

The Lagrange method can be applied to multiple constraints. In the mathematical exposition in Appendix A.2, *PVTTC* is made a continuous function of annual maintenance spending, $c_t$, in each year, *t*, from one to infinity. Annual budget constraints are imposed for years 1 to *m*. Thereafter, spending is unconstrained for years *m + 1* to infinity. The optimisation problem is

Minimise $PVTTC(c_1, c_2, \dots, c_m, c_{m+1}, \dots)$ subject to $c_t \leq B_t$ for all years *t = 1* to *m*

where $c_t$ is maintenance spending in year *t* and $B_t$ is the budget for year *t*.

The Lagrangian is

$$L = PVTTC(c_1, c_2, \dots, c_m, c_{m+1}, \dots) + \sum_{t=1}^{m} \lambda_t (c_t - B_t)$$

where $\lambda_t$ is the Lagrange multiplier for year *t*.

It is shown in Appendix A.2 that for any budget-constrained year *t* with a binding budget constraint, at the optimum

$$\lambda_t = -\frac{\partial PVTTC}{\partial c_t} = -\frac{dPVTTC}{dc_t}$$

The partial derivative, $\frac{\partial PVTTC}{\partial c_t}$, is the change in PVTTC from a change in spending in year *t* holding spending in all other years constant. The total derivative, $\frac{dPVTTC}{dc_t}$, is the change in PVTTC from a change in spending in year *t* with spending in all other years adjusted to minimise PVTTC. In the case of the total derivative, PVTTC is optimised both before and after the change in $c_t$. This equality is known generally as the envelope theorem (Cornes 1992).

In Figure 3.5, the value of $\lambda_1$ is $-\frac{\partial PVTTC}{\partial c_1}$ at point *D*.

An MBCR for a single year *t*, $MBCR_t$, can be defined as the saving in the present value of road user costs from increasing the budget in year *t* by one present-day dollar. The denominator of $MBCR_t$ has to be specified as a present value in year zero to be consistent with the BCR used for capital projects and the MBCR for present value budget constraints defined above. A small increase in maintenance spending in year *t* of $dc_t$, has a present value of $\frac{dc_t}{(1+r)^t}$, which is the denominator of $MBCR_t$.

The numerator of $MBCR_t$ is not just the saving to users, *dPVUC*, but also includes the present value of changes in road agency costs in years other than *t*, that is, $dPVAC - \frac{dc_t}{(1+r)^t}$. It is necessary to subtract the present value of the increase in maintenance spending in year *t* because it is already included in *dPVAC*.

Hence,

$$MBCR_t = -\frac{dPVUC + dPVAC - \frac{dc_t}{(1+r)^t}}{\frac{dc_t}{(1+r)^t}}$$

$$= -(1+r)^t \frac{dPVTTC}{dc_t} + 1$$

$$= (1+r)^t \lambda_t + 1 \qquad (3.9)$$

Thus, the MBCR for increasing the budget in year *t* can be obtained from the Lagrange multiplier for that year. The reason $\lambda_t$ has to be multiplied by $(1 + r)^t$ is that $\lambda_t$ is the saving in PVTTC from an additional dollar spent in year *t*. A one dollar increase in present day dollars equates to an increase of $(1 + r)^t$ year *t* dollars.

For unconstrained years and for years when the budget constraint is non-binding, $\lambda_t = 0$ and $MBCR_t = 1$, which is to be expected. The result in equation 3.9 is used in the case study in Chapter 6 to obtain annual MBCRs from the optimisation process.

In mathematical economics, Lagrange multipliers are 'shadow prices' because they trade the value of relaxing a constraint against the return in terms of the objective function (Bellman 1961, p. 103). Imagine that each road segment was controlled by a separate entity that incurs PVTTC in using and maintaining the segment. The entities have to bid for maintenance dollars to spend on their segment in specified years at a competitive auction. The values of the Lagrange multipliers would be the prices that would emerge from the auction expressed in year *t* dollars. Each segment owner would be willing to pay no more than the saving in PVTTC to them from obtaining an additional maintenance dollar. The market-clearing prices would be those that equate the demand for spending with the supply of funds in each of the budget-constrained years.

Imposing uniform budget constraints, as shown in the Chapter 6 case study, typically requires high $\lambda_t$ values in the early years when the demand for funds is high relative to budgets in order to suppress spending and push less economically warranted maintenance works into later years when funds are less scarce relative to demand and $\lambda_t$ values can be lower. Hence, with uniform annual budget constraints, annual MBCRs will be highest in the first year of the analysis period and will progressively fall as the backlog is caught up.

## 3.4.5 Incremental BCR for annual budget constraints

The MBCR for a single year *t*, $MBCR_t$, has been defined as the saving in the present value of road user costs from increasing the budget in year *t* by one present-day dollar. The IBCR equivalent, for a budget increase of any amount in a single year with a binding budget constraint is

$$IBCR = -\frac{\Delta PVUC + \Delta PVAC - \Delta PVB}{\Delta PVB} = -\frac{\Delta PVTTC}{\Delta PVB} + 1$$

where $\Delta PVB$ is the present value of the budget increase. $\Delta PVB$ could also be a present value for increases in multiple budget-constrained years. As with $MBCR_t$, it is necessary to subtract the present value of the budget increase from the numerator because it is already included in $\Delta PVAC$. After subtracting $\Delta PVB$ from $\Delta PVAC$, in the numerator, there remains other changes in PVAC due to changes in spending in years without budget constraints made by the optimisation process as a consequence of the increase in budgets for one or more years.[14]

## 3.4.6 Use of the MBCR for maintenance spending

For capital projects in a budget-constrained situation, the economically optimal choice of projects is found by selecting projects in descending order of BCR until either funds are exhausted or there are no more projects with BCRs above one, in which case the budget constraint is non-binding. The BCR of the last project to be accepted is the 'cut-off BCR' — the minimum acceptable BCR for a project to be funded.

If investment projects were finely divisible, the cut-off BCR could be termed the MBCR for capital expenditure. It indicates the benefit to society, expressed as a present value, of increasing the capital budget by an additional dollar. It could be compared with the MBCR for maintenance spending to see if the split of funds

----------

[14] The IBCR for increases in annual budget constraints could alternatively be defined as in Section 3.4.3 above for present value budget constraints with *ΔPVAC* in the denominator. The IBCR would then measure the saving in PVUC per dollar of change in PVAC instead of PVB. The alternative definition of the IBCR for relaxing annual budget constraints with *ΔPVAC* in the denominator was tested using the case study in Chapter 6. The results reported footnote 30 in Chapter 6 showed that the IBCR with *ΔPVAC* in the denominator can differ greatly from the IBCR with *ΔPVB* in the denominator and can behave in unexpected ways as budget constraints are relaxed. This arises from the finding in Chapter 6 that modest annual budget constraints in the early years of the analysis period reduce PVAC, but then increase it as the constraints are progressively tightened.

between investment and maintenance spending is economically optimal. Given a road agency budget to divide up between capital and maintenance spending, the optimal split would occur where the MBCRs for capital and maintenance spending are the same. If the MBCR for maintenance spending is above that for investment spending, economic welfare could be improved by shifting funds from the investment budget to the maintenance budget, and conversely. To illustrate, if the funding split was such that the MBCRs were 2.0 for capital and 4.0 for maintenance, shifting one dollar out of the capital budget would result in a loss of $2 in benefit, but added to the maintenance budget, the one dollar would earn a benefit of $4, a net gain of $2.

The MBCR also provides information about the absolute level of funding. In the absence of any costs of raising funds, all investment and maintenance spending with a BCR above one is warranted because it will lead to a net economic gain. If there is a cost of raising public funds, for example, the disincentive effects of increasing income taxes, a cut-off BCR might be set above one by an amount equal to the marginal cost of public funds. For example, if raising an additional dollar of tax to spend on infrastructure costs $0.30, any funds spent on capital projects or maintenance with a BCR below 1.3 would result in a net loss to the economy.

## 3.5 Cost-effectiveness analysis approach

An alternative approach to road maintenance optimisation common in the literature and used by road agencies is to minimise PVAC subject to minimum road condition constraints, for example, maximum allowable roughness levels. This is a form of cost-effectiveness analysis because it seeks to find the least-cost way to achieve the objectives of the specified minimum standards. Minimum road condition constraints are essential because without them, the unconstrained minimum value of PVAC would be zero in theory, or the lowest cost consistent with the technical constraints within the model in practice. Present value budget constraints are not relevant, but annual budget constraints can be imposed to smooth and defer road agency spending, albeit at the expense of a higher PVAC value.

The cost-effectiveness approach does not require a relationship between pavement condition and user costs in the model, but road users' interests would be considered when setting the condition constraints. Different road condition standards will be required for groups of roads with different traffic levels, vehicle mixes and locations. It is desirable that the condition standards follow a similar pattern to economically optimal standards. For example, roads with higher traffic levels and higher proportions of heavy vehicles should be maintained to higher standards. Road agencies may divide their networks into sub-networks for the purpose of applying appropriate maintenance standards, grouping together roads according to traffic and economic importance. Multiple iterations may be needed to find the set of standards that, according to subjective judgement, offers the 'right' distribution of standards across sub-networks and regions and fits within the budget.

Setting the minimum acceptable road conditions exogenously will almost certainly lead to a less economically efficient outcome compared to PVTTC minimisation, which allows the analysis to determine the maintenance standards considering users' costs.

Figure 3.7 illustrates the approach using the same two-period model as in Figure 3.5. Roughness in year one ($R_1$) is a decreasing function of agency spending in year one ($c_1$) and roughness in year two ($R_2$) is decreasing function of spending in both years 1 and 2 ($c_1, c_2$). The greyed area represents all combinations of spending in the two years for which roughness in one or both two years is below the acceptable maximum ($R_{max}$). The frontier along the south-west is the minimum value of $c_2$ for each given value of $c_1$ to remain within the zone. Points along the frontier just meet the maximum roughness standard and points to the left of and below the frontier are below the standard for one or both years.[15]

To minimise $PVAC = \frac{c_1}{(1+r)} + \frac{c_2}{(1+r)^2}$, in the absence of any annual budget constraint, it is necessary to find the line $p = \frac{c_1}{(1+r)} + \frac{c_2}{(1+r)^2}$ or $c_2 = p(1+r)^2 - c_1(1+r)$, with the smallest value of *p* consistent with values of $c_1$

----------

[15] For a single segment, the maximum roughness frontier, where $Max[R_1(c_1), R_2(c_1, c_2)] = R_{max}$ would be kinked at the point where the maximum roughness switches from the year-one roughness, $R_1$, to the year-two roughness, $R_2$, as $c_1$ is increased and $c_2$ reduced. For a group of segments considered together, the frontier would be approximately smooth.

and $c_2$ within the acceptable zone. This occurs where the line is tangent to the frontier of the acceptable zone at a point, *A*, with spending of $C_{1A}$ and $C_{2A}$.

Since the frontier is a level curve with a single maximum allowable roughness value,

$$dR = \frac{\partial R}{\partial c_1} dc_1 + \frac{\partial R}{\partial c_t} dc_t = 0$$

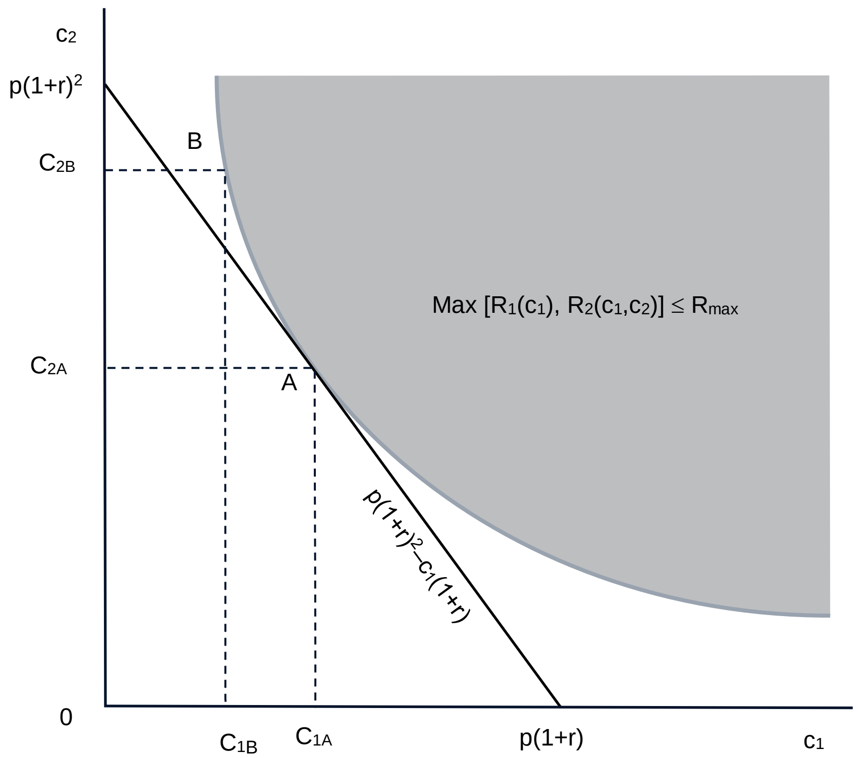
the slope of the frontier and the line at the point of tangency is

$$\frac{dc_2}{dc_1} = -\frac{\frac{\partial R}{\partial c_1}}{\frac{\partial R}{\partial c_2}} = -(1+r)$$

Higher discount rates will make the line steeper, moving point *A* north-west along the frontier, shifting spending from year one to year two.

Once an annual budget constraint is introduced, such as $c_1$ being constrained to the value $C_{1B}$, in Figure 3.7, the minimum PVAC is found at point *B* where $c_2$ is at a minimum consistent with the condition constraint region. However, unlike the case of Figure 3.5, point *B* is not a point of tangency with a level curve.

**Figure 3.7 Minimum PVAC subject to condition and budget constraints**




Mathematically, the problem is to minimise $PVAC = \sum_{t=1}^{\infty} \frac{c_t}{(1+r)^t}$ subject to a condition constraint that $Max[R_1(c_1), R_2(c_1, c_2), R_3(c_1, c_2, c_3), \ldots] \leq R_{max}$ and annual budget constraints $c_t \leq B_t$ for all years *t* = *1* to *m*. The condition constraint can be written more simply as $R(c_1, c_2, c_3, \ldots) \leq R_{max}$. Over an infinite time horizon, it is inevitable that the condition constraint will be binding because a road section cannot go without a roughness-reducing treatment indefinitely. The constraint then becomes one of equality, $R(c_1, c_2, c_3, \ldots) = R_{max}$.

Making this assumption, the optimisation problem is to minimise *PVAC* subject to $R(c_1, c_2, c_3, \ldots) = R_{max}$, and $c_t \leq B_t$ for all *t* = *1* to *m*.

The Lagrangian to minimise is

$$L = \sum_{t=1}^{\infty} \frac{c_t}{(1+r)^t} + \mu[R(c_1, c_2, \dots, c_t, \dots) - R_{max}] + \sum_{t=1}^{m} \lambda_t (c_t - B_t)$$

where $\mu$ is the Lagrange multiplier for the condition constraint. It is shown in Appendix A.3 that for any budget-constrained year $t$,

$$\lambda_t = -\frac{dPVAC}{dc_t}$$

A one dollar increase in the budget constraint for year *t*, where the constraint is binding, causes changes in required spending in other years leading to a gross change in PVAC of $\lambda_t$. It is a gross change because *dPVAC* includes the increase in spending in year *t*. The ratio of the net change in PVAC to $dc_t$, discounted to the present, is an MBCR indicating the net financial benefit to the road agency from spending an additional $(1+r)^t$ dollars in year *t* (equivalent to one dollar in year zero) not counting the increase in PVAC from the extra dollar. The subscript *a* is used to indicate that the benefits are limited to financial gains to the road agency.

$$MBCR_{at} = -\frac{dPVAC - \frac{dc_t}{(1+r)^t}}{\frac{dc_t}{(1+r)^t}} = (1+r)^t \lambda_t + 1$$

We could define an $IBCR_a$ to indicate the net benefit to the road agency from relaxing one more annual budget constraints by a non-small amount.

$$IBCR_a = -\frac{\Delta PVAC - \Delta PVB}{\Delta PVB}$$

## 3.6 Analysis period

The simple example in Section 3.3 had an infinite analysis period. Practical applications require a finite period. Switching from an infinite to a finite analysis period can alter the timings of maintenance treatments in the near future, which is of most interest, because the optimal time for a rehabilitation is affected by the costs and timings of future rehabilitations until such time as discounting makes the effect negligible. The effect is likely to be strongest when there are treatments due in the later years of the analysis period. With a finite analysis period, a model can save agency costs by deferring treatments to beyond the final year. At the end of the analysis period, the pavement is then left in poor condition. Treatments earlier in the analysis period might be delayed to compensate in part.

One solution is to extend the analysis period sufficiently far beyond the period of interest to make any such effects small. The lower the discount rate, the longer the necessary extension into the future. An example from the literature is Tsunokawa and Ul-Islam (2003, p. 197). They used a 40-year evaluation period to avoid the effects of employing an arbitrary assumption regarding the salvage value at the end of the analysis period.

An infinite time horizon, as in our simple model, is another solution. This is acceptable where a pavement can be assumed to reach a steady state with a uniform cycle of treatments into the indefinite future. Several models in the literature do this (see Table 4.1 in Chapter 4).

A simple and practical solution is to impose a minimum pavement condition constraint at the end of analysis period. For example, the constraint could be that the final condition be no worse than the initial condition (Ouyang and Madanat 2004, p. 355 and 2006, p. 769). However, an arbitrary final condition constraint could have significant effects on the selection and timings of treatments over the analysis period if the model over-maintains to meet a constraint set too high, or under-maintains to take advantage of a constraint set too low. Some models have minimum pavement condition constraints over the whole analysis period (see Table 4.1 in Chapter 4). This too can distort results if the constraint is set too high or too low.

Another practical solution is to minimise PVTTC over a limited number of years minus a 'residual value' or 'salvage value' of the road asset at the end of the analysis period. To compensate for the distortion caused by absence of costs after the analysis period, the residual value has to mimic the behaviour of PVTTC for costs after the analysis period. Alternatively, a depreciation amount could be added at the end of the analysis period. The residual value and depreciation amounts are mirror images of one another in the sense that when one goes up by a certain amount, the other goes down by the same amount.

The residual value can be defined as the price or construction cost of a new asset minus depreciation, *V – D*. Depreciation is zero for a new asset and rises as the asset ages reaching a maximum when the asset is fully worn out. The maximum will be the cost of restoring the asset in fully worn-out condition to new, or the replacement cost minus scrap value. Letting $PVTTC_n$ represent the PVTTC estimated over the analysis period of *n* years, minimising $PVTTC_n - (V - D)$ is the same as minimising $PVTTC_n + D$ because *V* is a constant.

Ideally, the model would be indifferent between implementing a treatment at the end of the analysis period and not implementing it because the residual value or depreciation value changes by an offsetting amount.

PVTTC can be split into two parts, costs in years from zero to the end of the analysis period, $t^*$, and costs in years after $t^*$ to infinity, with the latter discounted back to year 0 by multiplying it by $e^{-rt^*}$. $\delta^*$ is the pavement age at the end of a finite analysis period.

$$PVTTC(T,\delta) = PVTTC(T,\delta)|_0^{t^*} + e^{-rt^*} PVTTC(T,\delta^*)|_{t^*}^{\infty}$$

where

- $PVTTC(T,\delta)|_0^{t^*}$ = present value of total transport posts over the analysis period, years 0 to $t^*$ , with a pavement age at year zero of $\delta$.
- $PVTTC(T,\delta^*)|_{t^*}^{\infty}$ = present value of total transport costs after the analysis period discounted to year $t^*$.

To counter the effect of a finite analysis period, the residual value has to mimic the behaviour of $PVTTC(T,\delta^*)|_{t^*}^{\infty}$ as *T* and $\delta^*$ change. In the simple illustrative model in this chapter with a single rehabilitation treatment, PVTTC increases with pavement age in year zero. The range over which PVTTC changes over the range of pavement ages in year zero, $0 \le \delta \le T$, is the rehabilitation cost.

To demonstrate this, equations 3.1 and 3.2 above, which express PVTTC as a function of cycle length, *T*, and pavement age at year zero, $\delta$, can be written as

$$PVTTC(T,\delta) = \int_0^{T-\delta} u(t+\delta)e^{-rt}dt + \frac{e^{-r(T-\delta)}}{(1-e^{-rT})}\left[c + \int_0^T u(t)\, e^{-rt}dt\right]$$

The difference in PVTTC between an old pavement just prior to rehabilitation, $\delta = T$, and a new pavement, $\delta = 0$, is the rehabilitation cost, *c*.

$$PVTTC(T,T) - PVTTC(T,0) = \left[\frac{c+\int_0^T u(t)\, e^{-rt}dt}{(1-e^{-rT})}\right] - \left[\int_0^T u(t)e^{-rt}dt + \frac{e^{-rT}\left[c+\int_0^T u(t)\, e^{-rt}dt\right]}{(1-e^{-rT})}\right] = c \qquad (3.10)$$

Using our numerical example to illustrate the point, with cycle time, *T*, set equal to the optimal time of 40 years, the difference in *PVTTC* between a pavement of 40 years of age and a new pavement (zero years of age) is, $PVTTC(40, 40) - PVTTC(40, 0)$ = \$2.16 million – \$0.83 million = \$1.33 million, the cost of a rehabilitation. Note that equation 3.10 holds for all values of *T*, not just the optimal value.

With a residual value that mimics $PVTTC(T,\delta^*)|_{t^*}^{\infty}$, if the model attempts to reduce PVTTC by delaying the last rehabilitation until after the analysis period, which will save a cost of *c* dollars, the cost saving is exactly negated by the residual value falling (or depreciation rising) by *c* dollars.

Using our numerical example, Figure 3.8 shows that straight-line depreciation can be a reasonable approximation for an infinite time horizon.

Setting the value of a newly-rehabilitated pavement, *V*, equal to the rehabilitation cost, *c* = \$1.3 million, the three curves show

- residual value from straight-line depreciation: $[PVTTC(40,40) - PVTTC(40,0)]\left(1 - \frac{\delta^*}{40}\right) = c\left(1 - \frac{\delta^*}{40}\right)$
- exact residual value: $c + PVTTC(40,0) - PVTTC(40,\delta^*) = PVTTC(40,40) - PVTTC(40,\delta^*)$

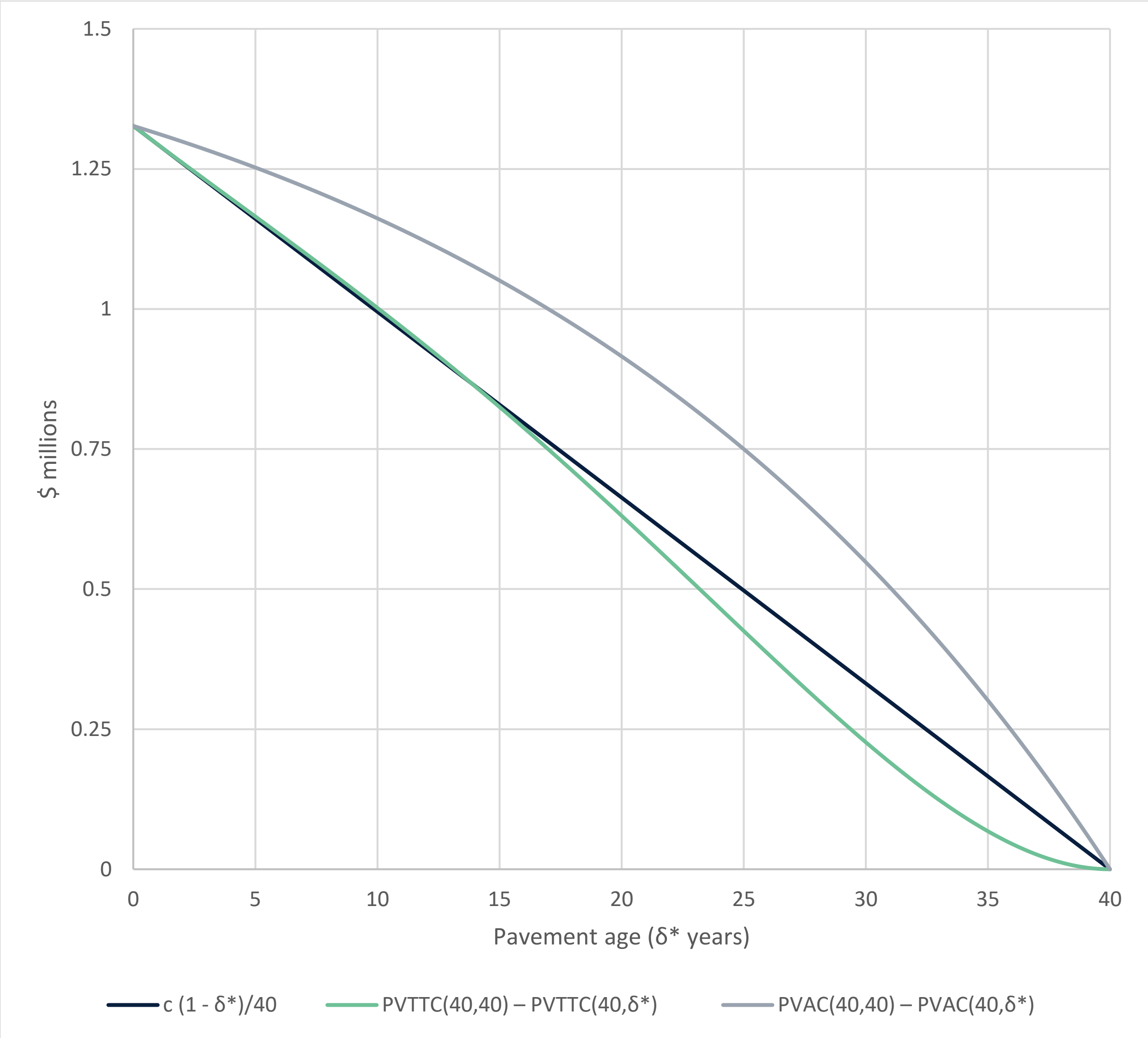
- exact residual value for road agency costs: $c + PVAC(40,0) - PVAC(40,\delta^*) = PVAC(40,40) - PVAC(40,\delta^*)$.

The residual value for road agency costs is relevant for minimising road agency costs subject to pavement conditions as discussed in Section 3.5. From equation 3.5, $PVAC = \frac{e^{-r(T-\delta)}}{(1-e^{-rT})}c$.

$$PVAC(T,T) - PVAC(T,0) = \frac{1}{(1-e^{-rT})}c - \frac{e^{-rT}}{(1-e^{-rT})}c = c$$

showing that, just as for PVTTC, the range of variation in PVAC over the course of the cycle is the rehabilitation cost, *c*.

**Figure 3.8 Residual value decline as a function of pavement age: actual and straight-line depreciation**




Straight line depreciation will be more approximate when the pavement is part-way through the cycle, when it differs most from the actual *PVTTC* or *PVAC*, as Figure 3.9 illustrates. Another source of approximation arises from the assumptions in the simple illustrative model of uniform cycles and a single treatment type. Changes in traffic levels and vehicle mixes over time will cause optimal cycle times to change over time. Different treatment types will restore a pavement by different amounts and have different costs. A practical solution to the problem of different treatment types is to use condition-based depreciation, for example, based on roughness or a combination of condition measures.

Due to the approximate nature of the depreciation estimate in its role of serving as a proxy for *PVTTC* or *PVAC* beyond the analysis period, it is still advisable to have an analysis period that extends well beyond the period of interest so as to reduce the impact of errors in the depreciation approximation. Our case study below uses an analysis period of 40 years, while budget constraints are imposed only for the first 10 or 20 years. Sensitivity tests are undertaken of shortening the analysis period to 30 and 20 years to show the effects of greater reliance on the depreciation estimate.

## 3.7 Multiple treatment types

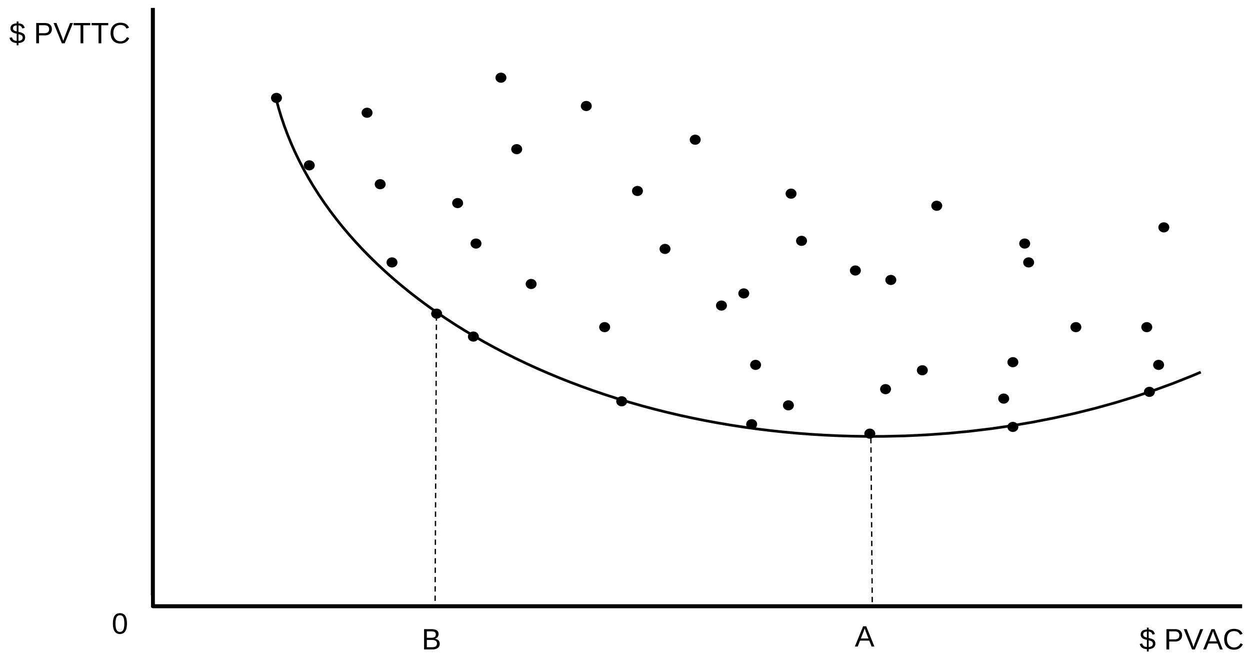
Multiple treatment types can be modelled either by making treatment intensity a continuous variable or having the model choose between alternative discrete treatment types. A more intense treatment, such as a greater overlay thickness, will have a greater effect on reducing roughness but will cost more to implement.

In Chapter 4, Section 4.3.1.1, it is reported that models with continuous pavement intensity almost invariably find that the optimal strategy is to resurface either to the highest or the lowest standard permitted by the model (a corner solution) rather than an intermediate standard that balances the marginal benefit and marginal cost of increasing overlay thickness.

With multiple discrete treatment types to choose between, the optimisation problem becomes much more complex. Instead of a smooth, continuous cost surface with a single minimum point for a single segment, there are discrete choices between treatment types giving rise to multiple local minimums.

Figure 3.9, based on a diagram in Tsunokawa and Ul-Islam (2003, p. 196), shows how the results might appear for a single segment if PVTTC values from various options with multiple treatment types were plotted against the present value of maintenance costs. The term 'option' is used throughout this report to refer to a set of treatments, distinguished by type and implementation time, to be carried out over the analysis period. For each spending level (PVAC), only the options with the lowest PVTTCs are of interest. Joining the minimum values together with a smooth curve would produce a U-shaped curve, the same relationship as shown in Figure 3.3. The minimum point of the curve, at spending level *A*, is the unconstrained optimum. The point on the curve above spending level *B* is shows the optimum with a present value budget constraint of *B*.

**Figure 3.9 Illustration of present values of total transport costs from different maintenance options**




## 3.8 Optimising the investment–maintenance trade-off

While the present report focusses on maintenance with design standards as taken given, a short digression is included here on the economic principles of optimising the trade-off between investing in stronger, more durable pavements to save on maintenance costs.

Incurring higher investment costs to construct a stronger pavement at the outset saves future maintenance and user costs. For flexible pavements, greater initial pavement strength (a higher structural number) leads to a lower deterioration rate, as Paterson's algorithm, presented in Chapter 2, illustrates. In the extreme, a concrete pavement costs much more than a flexible pavement to construct but requires far less future maintenance spending to provide a given level of service to users.

The optimisation problem in the absence of budget constraints has the same form as Figure 3.2 above. With pavement strength on the horizontal axis, as pavement strength increases, construction costs rise and the optimised PVTTC for maintenance falls. By 'optimised PVTTC' here, is meant the PVTTC for maintenance for each pavement strength that minimises the sum of road agency and user costs given pavement strength. Vertically adding the upward-sloping capital cost curve and downward-sloping maintenance cost curve produces a U-shaped total cost curve. Letting *K* represent capital costs and *s*, pavement strength, the optimum occurs where

$$\frac{\partial K}{\partial s} = -\frac{\partial PVTTC}{\partial s}$$

that is, where the additional capital cost from increasing pavement strength by one unit equals the resultant saving in PVTTC. The optimum pavement strength condition can be rewritten as $-\partial PVTTC/\partial K = 1$. The interpretation is that investment in increased pavement strength should occur as long as the benefit from each in additional dollar invested, that is the saving in PVTTC, exceeds one. As more is invested in pavement strength, the law of diminishing returns causes the marginal benefit to fall. It is not worthwhile to invest beyond the point where the marginal benefit falls below the one-dollar marginal cost. Higher traffic levels are associated with stronger pavements because the savings in user costs from additional spending on stronger pavements are greater.

There is little literature on optimising the pavement strength–maintenance trade-off. Tsunokawa and Ul-lslam (2003) tested combinations of initial pavement design options with varying strengths, maintenance options (condition-responsive overlays with varying threshold roughness values and thicknesses), traffic loadings, and national economic characteristics (discount rates and wage rates, which affect the value of time for users and agency costs).

According to the above formulation of the optimum pavement strength condition, the benefits from a stronger pavement are realised as a combination of savings in user costs and maintenance costs. In the models of Small et al. (1989) and Newbery (1989), the benefit from a stronger pavement is realised entirely in the form of a saving in maintenance costs, with no change to user costs. In their models, the intervention roughness level is exogenous, so a stronger pavement increases the time interval between rehabilitations with no impact on the present value of user costs.[16] With user costs fixed, their models minimise *K + PVAC*. The resulting optimal condition is that pavement strength should be adjusted to set $\partial K/\partial s = -\partial PVAC/\partial s$. The marginal investment cost from building a slightly stronger pavement is equated with the marginal benefit of a reduced maintenance cost to the road agency.

We now consider optimal pavement strength in the presence of budget constraints on maintenance and capital spending.

----------

[16] Newbery's (1989) model has an interesting conceptual simplification. For a single homogeneous segment of pavement considered in isolation, any change in rehabilitation timings would change the present value of user costs, but for a group of identical segments with a uniform age distribution taken together, there is no change. A vehicle travelling over all the segments would experience the full range of possible roughness levels. For example, with a 40-year rehabilitation period, there would be 40 one-kilometre segments with end-of-year ages from 1 to 40. At the end of each year, the oldest kilometre of pavement is rehabilitated, so the average age for the 40 segments does not change from year to year. If the rehabilitation period was raised to 60 years as a result of building a stronger pavement, there would be 60 segments of two-thirds of a kilometre in length, with end-of-year ages from 1 to 60. The average roughness, and hence user cost, would be the same in both cases, but the agency's annual rehabilitation cost would be one third lower than for the 40-year rehabilitation period.

If pavement strength for a segment is increased by one unit, the cost will be $\partial K/\partial s$. If the cost comes out the capital budget for which the cut-off BCR is $MBCR_k$, the opportunity cost of the unit increase in pavement strength is the benefits forgone by not investing the capital cost in infrastructure projects, $\frac{\partial K}{\partial s}MBCR_k$.

The unit increase in pavement strength for a particular segment means that $\partial PVAC/\partial s$ in maintenance costs can be saved with no reduction in user costs for that segment. That saving effectively increases the maintenance budget yielding a benefit of $MBCR_m = -\partial PVUC/\partial PVAC$. The benefit from a unit increase in pavement strength is therefore $-\frac{\partial PVAC}{\partial s}MBCR_m$. The negative sign is needed because PVAC reduces as pavement strength increases.

The optimum pavement strength occurs where there are no longer any gains to made by shifting funds between spending on capital projects and investing in pavement strength, that is, where

$$\frac{\partial K}{\partial s}MBCR_k = -\frac{\partial PVAC}{\partial s}MBCR_m$$

This can be rewritten as

$$-\frac{\partial PVAC}{\partial K} = \frac{MBCR_k}{MBCR_m} \tag{3.11}$$

The last expression implies that the optimal pavement strength is found where the saving in PVAC from an additional dollar spent on building a stronger pavement is equated to the ratio of the MBCRs for capital and maintenance spending. The more maintenance spending is constrained relative to capital spending, the lower the ratio $MBCR_k/MBCR_m$. If maintenance spending is more constrained than capital spending, there will be net economic benefits from diverting funds within the capital budget away from new infrastructure projects to build stronger pavements reducing the pressure on the more constrained maintenance budget. The converse applies if capital spending is more constrained than maintenance spending.

Using HDM-4 simulations, Tsunokawa and Ul-Islam (2003) showed that optimal pavement strength is higher with budget-constrained maintenance spending. They made an implicit assumption that investment funds are unconstrained ($MBCR_k = 1$). Their finding is consistent with equation 3.11, because, if the with right side of equation 3.11 is below one due to $MBCR_m > 1$, more has to be spent on pavement strength in order to bring the left side of the equation down below one.

It is clearly a 'second-best' outcome to have the pavement strength decision affected by different relative scarcities of funds in separate investment and maintenance budgets. It would be better to ensure similar MBCRs for investment and maintenance, even if both are above one. Then optimal pavement strength will be found where $-\partial PVAC/\partial K = 1$, that is, increase strength up to the point where the saving in PVAC from the marginal dollar invested equals one.

## 3.9 Optimal incentives in maintenance contracts

The model of maintenance optimisation developed here fits neatly into the incentive regulation framework for commercial road supply in Harvey (2015). Harvey's framework applies to the complete supply of road services by a public utility or private firm. It is shown here how it can be applied at a less ambitious level to outsourcing of maintenance activities to a contractor. This is often done with performance-based contracts in which the contractor is required to ensure a road meets specified minimum condition standards rather than being required to carry out specified works (Zeitlow 2006). Under the terms of the contract, financial penalties are imposed for failing to meet the required condition standards. In Harvey's scheme, the sole performance measure is generalised user costs and the penalties for under-performance and bonuses for over-performance exactly equal their marginal social values. The model and parameters for estimating user costs form part of the performance contract. With the impact of road condition on user costs effectively internalised into the supplier's revenue stream, profit maximising behaviour by the supplier leads to the welfare maximising outcome.

For simplicity, we initially assume an infinite time horizon. It is also assumed that the social and private discount rates are identical. Adapting Harvey's shadow toll formula, for a single road segment, the maintenance contractor is paid in year *t* of the maintenance cycle, an annual sum of $p - [c(t) - c(0)]$ where *p* is a constant, *c(t)* is annual user cost in year *t* and *c(0)* is annual user cost with a new pavement.

The annual payment would then be highest just after the pavement had been rehabilitated, at which time $c(t) - c(0) = 0$. Payments would reduce as the pavement becomes rougher causing *c(t)* to rise above *c(0)*. Thus annual maintenance payments follow a sawtooth pattern over time, the mirror image of the pattern for user costs. With *r* the discount rate and an infinite time horizon, the present value of the contractor's profit, *Π*, is

$$PV\Pi = \frac{p}{r} - \left[PVUC - \frac{c(0)}{r}\right] - PVMC = \frac{[p - c(0)]}{r} - PVTTC$$

where *PVMC* is the present value of maintenance costs incurred by the contractor (equivalent to PVAC).

As defined earlier in this chapter, PVUC and PVMC are functions of the time interval between treatments, *T*. To maximise profits, the contractor would schedule maintenance treatments to set *dPVΠ/dT = 0*. Since $[p - c(0)]/r$ is constant with respect to maintenance expenditure, the contractor will aim to set *dPVTTC/dT = 0*, which is the same as the condition for minimising social costs. A competitive tender process would ensure that *p* was bid down to the point where $[p - c(0)]/r - PVTTC = 0$ with *r* including a normal return on investment. The value of *c*(0) is exogenous. A different value of *c*(0) would be exactly offset by a changed value of *p*. The value of *p* would be affected by pavement age at the start of the contract, being higher for an old pavement because future rehabilitations occur sooner. With the reward for over-performance and the penalty for under-performance exactly mirroring the impacts of pavement condition on user costs, the contractor has then an incentive to maintain at the optimal standard.

Harvey's shadow toll formula includes a 'correction factor', $\psi$ that modifies the incentive faced by the supplier in situations when the regulatory authority or road agency wishes to deliberately engender over- or under-investment or maintenance. This might occur for maintenance if the road agency had constrained funds necessitating underspending, or had to meet an obligation imposed by the government to provide a minimum service level requiring overspending. The annual payment to the maintenance contractor becomes, $p - \psi[c(t) - c(0)]$ and the contractor's profit function is

$$PV\Pi = \frac{p}{r} - \psi\left[PVUC - \frac{c(0)}{r}\right] - PVMC = \frac{[p - \psi c(0)]}{r} - \psi PVUC - PVMC$$

Profit maximisation requires the contractor to set

$$\frac{dPV\Pi}{dT} = -\psi\frac{dPVUC}{dT} - \frac{dPVMC}{dT} = 0 \quad \text{which implies} \quad \psi = -\frac{dPVMC}{dPVUC} = -\frac{1}{MBCR}$$

If the road had be maintained at a standard below the economic optimum for budgetary reasons, (MBCR > 1), the road agency would set $\psi$ below one at the reciprocal of the MBCR that achieves the budget constraint. This reduction in the financial reward for better maintenance leads to a lower standard of maintenance being provided. The tender process would ensure that the value of *p* was lower than otherwise so the contactor's costs were exactly covered. Conversely, if the aim was to maintain the road at an above-optimal standard, the road agency would set $\psi$ above one.

If the contract was to terminate at a given future date, the contractor faces an incentive to defer the last treatment. The incentive could be removed by including in the contract an end-of-term adjustment whereby one party pays the other the difference between the actual depreciation amount and an agreed depreciation amount.

## 3.10 Conclusion

The underlying principle of road maintenance economics is to minimise the present value of total transport costs, trading off the benefit of savings in user costs from higher maintenance standards against the additional costs to the road agency. For the simplest case where there is a single treatment type, an optimisation model has to find the optimal times to implement treatments. For a more complicated model, there will also be multiple treatment options to choose between. Budget constraints can be introduced expressed either as a present value of road agency costs or as annual spending maximums. The latter is more realistic but, as subsequent chapters will demonstrate, is much more difficult analytically.

The marginal and incremental benefit–cost ratio (MBCR and IBCR) concepts introduced in this chapter offer decision makers useful information about the economic value of adjusting maintenance spending and can be compared with BCRs for investment projects. Using the formula derived in Section 3.4.4, $MBCR_t = (1+r)^t \lambda_t + 1$, values for MBCRs can be obtained from the optimisation process, which is demonstrated by the case studies in Chapters 5 and 6.

Having a finite analysis period in maintenance modelling can distort the optimal solution. Ways to minimise the distortion include subtracting a residual value or adding a depreciation amount at the end of the analysis period, and setting the analysis period well beyond the period of interest.

The topics of optimising the pavement strength and setting optimal incentives in maintenance contracts were addressed at a theoretical level at the end of the chapter, drawing on the concepts and model developed. These are not pursued further in the remainder of the report.

# 4. Maintenance optimisation modelling literature review

## Summary

There is large body of literature from the civil engineering discipline on road maintenance optimisation in which authors specify a problem and present one or more solution methodologies. In most cases, each article uses a quantitative case study to illustrate and test the methodology. In general terms, the problem is to select a set of maintenance treatments from a menu of alternative treatment types, each with its own effects on pavement condition and implementation costs, together with times to implement the selected treatments, that maximises or minimises an objective function subject to budget and/or pavement condition constraints.

Some studies minimise the present value of total transport costs without budget constraints, which yields the most economically efficient solution. A number do this with present value, annual or average annual budget constraints. Some minimise either the present value or undiscounted sum of road agency costs over the analysis period subject to minimum pavement condition constraints. Then there are models that maximise a variety of pavement condition measures.

Most of the deterioration models in the road maintenance optimisation literature are either deterministic with continuous pavement condition or probabilistic with discrete pavement condition adopting a Markov chain approach. There is disagreement in the literature about the advantages and disadvantages of deterministic and probabilistic approaches.

The number of possible solutions to road maintenance optimisation problems rises exponentially with the numbers of treatment types, analysis years and segments. This is known as the 'curse of dimensionality' or 'combinatorial explosion'. Each case study in the literature has to manage the curse of dimensionality through a combination of restricting the numbers of segments analysed together, treatment types and analysis years, and by applying a suitable optimisation method.

Some early articles used 'prioritisation' approaches which, while straightforward to implement, are unable to find genuine optimum solutions over the long term.

Over the last 40 years during which most of the literature has developed, increasing computer speeds have vastly increased the maximum feasible size of problems that can be optimised. Mathematical optimisation techniques such as linear, integer and dynamic programming are able to find optimum solutions provided the number of possible solutions is not too great. Heuristic optimisation algorithms, such as genetic algorithms, have extended the size of problem that can be accommodated. They can find good solutions but the solutions are not necessarily the overall optimum. Their effectiveness declines for extremely large problems.

Two-stage approaches have been developed in which the best or a number of better solutions are found for each segment considered in isolation without any budget constraints in the first stage, followed by a prioritisation approach employed in the second stage to choose a set of solutions that fits within annual budget constraints.

## 4.1 Introduction

This chapter reviews the road maintenance optimisation literature and discusses alternative approaches to road maintenance optimisation modelling. The literature on application of mathematical optimisation methods applied to road maintenance is voluminous, and this report makes no claim to complete coverage. The literature review concentrates on journal articles, with only a few conference papers cited. Some articles were not cited because they provided insufficient detail on their approach. There is also a substantial body of related literature not examined on maintenance optimisation in general and applied to infrastructure other than roads, such as bridges.

Only two of the articles cited in this chapter, Abelson and Flowerdew (1975) and Gerchak and Waters (1978), are from the economics discipline. The rest are from civil engineering journals and conferences. Consequently, a significant proportion of the articles adopt frameworks not aligned with economic principles. Examples include use of prioritisation methods that do not lead to optimal solutions, optimising non-monetary objectives such as a pavement condition indicator, and not discounting costs.

There are different ways to group the various approaches in the literature. In this literature review, models are split into those with continuous and discrete pavement condition and within each of these categories, deterministic and probabilistic models. Each of these four groups is discussed in turn. Then the chapter addresses some of the major technical issues faced when attempting to optimise road maintenance, in particular, the potentially huge number of combinations of treatment types and possible treatment times, sometimes referred to as the 'curse of dimensionality' or 'combinatorial explosion'. Others matters addressed include the variety of objective functions, constraints and optimisation techniques found in the literature.

## 4.2 Prioritisation methods

The simplest approach to maintenance management is to base decisions on current pavement condition. The worse the pavement condition, the higher the priority to treat the pavement (Zimmerman 1995, p. 5). Early pavement management systems ranked pavements according to a simple condition measure, such as a weighted index of current distresses. Treatment projects would be developed for the higher-ranking pavements, costs estimated, and projects chosen in descending order until the available budget was exhausted (Kulkarni and Miller 2003). Ideally, adjustments would be made to rankings for the class of road or traffic level, recognising that, for economic and community expectations reasons, more important roads in a network should be kept in better condition than less important roads. Some subjective pavement condition measures take account of the condition required for a road to meet its traffic task. For example, Shah et al. (2014) prioritised maintenance according to the product of four indicators: a road condition index, a traffic volume factor, a drainage factor and a factor related to the importance (class) of the road.

Wang et al. (2003, p. 22) pointed out two major disadvantages of ranking methods. First, with the most severely damaged pavement segments given the highest priority, the highly ranked segments can consume the entire budget, ignoring the effect on overall network condition. Second, treatment timing is not well handled. For example, some segments with a low ranking may deteriorate rapidly over the coming few years, necessitating a much more costly later treatment.

The next level approach to pavement management is to prioritise maintenance projects across a network and over time using a rule that takes account of benefits and costs. The most common approaches are cost-effectiveness (that is, improvement in a road condition measure divided by increase in cost) and benefit–cost ratios (BCRs), which may be defined in a variety of ways. The time period over which benefits are measured can be as short as just one year or much longer using a deterioration model.

One form of cost-effectiveness ratio is the area under a performance curve (that is, a pavement condition index as a function of time, with higher values of the index representing better condition) above a minimum acceptable level divided by the treatment cost (Zimmerman 1995, p.9; Li et al. 1998). When the index is below the minimum acceptable level, the area is counted as zero. The ratio is multiplied by average annual daily traffic (AADT). Under the 'priority programming' approach of Li et al. (1997b), the sum of cost-effectiveness ratios of treatment alternatives for a network over an analysis period was maximised using prioritisation to find the optimum.

Replacing the condition index with road user cost saving compared with road user cost at the minimum acceptable pavement condition, produces a similar curve that declines over time at an increasing rate following a treatment. Road user cost indicates users' willingness-to-pay for improvements in pavement condition and so is an improvement on condition indexes from an economic point of view. The resultant cost effectiveness ratio could be termed a benefit–cost ratio. The HDM-4 model prioritises treatments using an 'incremental net present value/cost ratio' defined as the difference between the net present value of the selected project alternative and the net present value of the base alternative, divided by the cost of the selected project (Odoki and Kerali 2006, Volume 4, A1-17 and G1-26; Part A,).

The main weakness of prioritisation methods is that they do not ensure the best possible maintenance strategies over long planning time spans (Meneses and Ferreira 2013; Jorge and Ferreria 2012). For alternatives that require more than one treatment during the budget-constrained years (for example, treatments in years 3 and 15), the incremental BCR approach would take account only the budgetary cost of the first treatment. Effects of any decision about the next treatment on future optimal treatment types and times are ignored.

## 4.3 Overview of optimisation models

Finding the optimum set of treatment types and times requires use of an optimisation model. The discussion in this chapter categorises models in the literature first according to whether pavement condition is treated as a continuous variable (for example, roughness or pavement condition index) or a discrete variable with a number of condition states specified. At the second level, models are categorised according to whether they are deterministic or probabilistic. Deterministic models forecast single values for one or more condition measures at each point in time while probabilistic models estimate multiple values at each point in time with probabilities attached to them.

Following this categorisation method, Tables 4.1 and 4.2 list the articles reviewed that apply optimisation techniques to road pavement maintenance. The articles in the tables all present a model, or models, illustrated by one or more case studies. In broad terms, the models seek to maximise or minimise an objective function usually subject to constraints. For one or more pavement segments or 'families' of segments grouped together, the model decides for each year of the analysis period which treatment type to apply from a menu of alternatives that includes the null treatment (do-nothing or do-minimum).

### 4.3.1 Continuous pavement condition group — Table 4.1

#### 4.3.1.1 Continuous pavement condition and deterministic category

The simple model presented in Chapter 3 is an example of a deterministic model with continuous pavement condition. It has time as a continuous variable and a single treatment type — rehabilitation. These characteristics are shared by some of the models in Table 4.1. Deterioration functions are usually specified as simple exponential curves that are a function of time. Models of this type are aimed at understanding the determinants of optimal maintenance policies and reaching general conclusions rather than serving as tools with which to develop actual maintenance programs.

Li and Madanat (2002, p.526) define the problem as, "given a known deterioration curve and rehabilitation effectiveness, what is the frequency and intensity of pavement rehabilitation activities that minimises the discounted social (agency plus user) cost over a long planning horizon". The rehabilitation time that is optimised is sometimes a single instance between time zero and a given time for the second rehabilitation, or a uniform time interval between rehabilitations over an infinite time horizon, or a series of treatment times. Models with single rehabilitation treatments and deterioration curves are unable to capture the effect of reseals on deterioration rates, illustrated in Figure 2.2 in Chapter 2.

**Table 4.1 Models with continuous pavement condition**

| Article | Objective function | Non-technical constraints | Optimisation method | Treatment types | Pavement condition | Analysis period (years) | Time | Segments / families |
|---|---|---|---|---|---|---|---|---|
| **Deterministic models** | | | | | | | | |
| **Mahoney et al. 1978** | max gains of pavement rating points times survival probabilities | undiscounted budget & resource | integer programming | 7 | PCR | 1 | disc | 15 |
| **Büttler & Shortreed 1978** | max benefit as a function of smoothness | none then budget per time period | optimal control | cont function | riding comfort index | variable | cont | 1 |
| **Friesz & Fernandez 1979** | max net benefits (effectively min PVTTC) | none | optimal control | 1 | PSI | variable | cont | 1 |
| **Markow & Balta 1985** | max net benefits (effectively min PVTTC) | none | optimal control | 1 | PSI | one rehab | cont | 1 |
| **Bhandari et al. 1987** | min PVTTC | none, then average annual budget | full enumeration | 30 policy alternatives | IRI | 25? | disc | na |
| **Hajek & Phang 1988** | max years in which pavement performance above acceptable level x segment length, then former weighted by traffic, then area under PCI curve | annual budget | linear programming & mixed integer linear programming | 3 | PCI | 5 | disc | 134 |
| **Fwa et al. 1994b** | min PVAC | annual budget & minimum condition | genetic algorithm | 3 | various | 20 | disc | 45 |
| **Tsunokawa & Schoffer 1994** | min PVTTC | none | optimal control | cont function | QI | infinite | cont | 1 |
| **Fwa et al. 1996** | min PVAC | minimum condition | genetic algorithm | 9 | PSI | 20 | disc | 20 |
| **Pilson et al. 1999** | max best performance indicator & min undiscounted agency cost ** | none | genetic algorithm | 3 | 0-1 scale | 25 | disc | 1 |

| Article | Objective function | Non-technical constraints | Optimisation method | Treatment types | Pavement condition | Analysis period (years) | Time | Segments / families |
|---|---|---|---|---|---|---|---|---|
| **Abaza 2002** | min 'PLC disutility' (ratio of PLC cost including initial construction, then max maintenance & rehab to area under PLC performance curve) | none | trial & error for 1 to 6 rehabs over analysis period | 1 | PSI | variable (integer number of cycles) | cont | 1 |
| **Li & Madanat 2002** | min PVTTC | none | differentiation | 1 | QI | one rehab | cont | 1 |
| **Tsunokawa & Ul-Islam 2003** | min PVTTC | none | full enumeration | 8 overlay thickness options | IRI | 40 | disc | 288 *** |
| **Wang et al. 2003** | max treatment effectiveness & min disturbance to road users ** | annual budget & minimum condition | mixed integer programming | 5 | score | 5 | disc | 10 |
| **Ouyang & Madanat 2004** | min PVTTC | average annual budget | mixed integer nonlinear programming & greedy | 1 | QI | 60 | disc | 3 |
| **Ouyang & Madanat 2006** | min PVTTC | none | authors' method | 1 | QI | 100 | cont | 1 |
| **Tsunokawa et al. 2006** | min PVTTC | none | gradient method | variable overlay thickness | IRI | 20 | disc | 1 |
| **Ul-Islam & Tsunokawa 2006** | min PVTTC | none | gradient method | variable overlay thickness | IRI | 25 | disc | 1 |
| **Scheinberg & Anastasopoulos 2010** | min undiscounted agency cost or max wtg avg condition or max % of network > threshold condition | Undiscounted budget &/or wtd avg condition &/or % network > condition threshold | mixed integer programming | 8 | critical condition index &/or IRI | 3 | disc | 4 |
| **Sathaye & Madanat 2011** | min PVTTC | undiscounted budget | differentiation & greedy | 1 | QI | one rehab | cont | 3 |

| Article | Objective function | Non-technical constraints | Optimisation method | Treatment types | Pavement condition | Analysis period (years) | Time | Segments / families |
|---|---|---|---|---|---|---|---|---|
| **Gu et al. 2012** | min PVTTC | none | numerical methods | 2 | QI | infinite | cont | 1 |
| **Jorge & Ferreira 2012** | min PVTTC | annual budget | not specificied | 8 | PSI | 20 | disc | 67 |
| **Rashid & Tsunokawa 2012** | min PVTTC | none | optimal control | 3 cont | IRI | infinite | cont | 1 |
| **Meneses & Ferreira 2013** | min undiscounted agency & min user costs** | none | genetic algorithm | 4 | PSI | 20 | disc | 32 |
| **Torres-Machi et al. 2014** | max PSI units (area under curve) | none then with PV budget | simulated annealing then effectiveness/cost ratio * | 6 | PSI | 25 | disc | 5 |
| **Chen et al. 2015** | max PV benefit & min PVAC ** | acceptable pavement performance index & annual budget | dichotomic algorithm & GA | not specified | pavement performance index | 10 | disc | 699 |
| **Yepes et al. 2016** | max area under PCI curve | PV budget & annual budget | greedy then threshold accepting * | 12 asphalt, 6 concrete | PCI | 25 | disc | 20 |
| **Probabilistic models** | | | | | | | | |
| **Jawad & Ozbay 2006** | min PVTTC | none | genetic algorithm | 3 | IRI | 30? | disc | 20 |
| **Gao & Zhang 2008** | min undiscounted AC & min prob of exceeding upper IRI bound ** | upper bound on IRI | unspecified linear | 4 | IRI | 5 | disc | 1 |
| **Ng et al. 2011** | min undiscounted AC | minimum acceptable condition scores for pavement segments | integer programming | 3 | integer condition score | 5 & 10 | disc | 351 |

Notes: * in the optimisation method column indicates a two-stage optimisation process
The numbers in the Treatment types column exclude the null treatment or routine maintenance treatment alone.
** in the objective function column means bi-objective
*** 4 economies x 3 traffics x 20 pavement designs =288
Abbreviations: max = maximise, min = minimise, PV = present value, TTC = total transport cost, AC = average cost, wtd = weighted, avg = average, prob = probability, cont = continuous, disc = discrete, rehab = rehabilitation, QI = quarter index, PSI = pavement serviceability index, PCI = pavement condition index, PLC = pavement life cycle

The optimisation technique employed by the earliest articles in the continuous pavement condition deterministic category was optimal control theory, until it was realised that simpler methods would suffice. Friesz and Fernandez (1979) specified maintenance spending as an ongoing amount over time, which may be discontinuous (a 'bang-bang' control policy that jumps between levels).[17] Markow and Balta (1985) used optimal control theory to solve for optimal timing for a single rehabilitation event in a finite time period. Tsunokowa and Schofer (1994, p. 153) observed that optimal control models with discrete jumps in the state variable, that is, the pavement condition improvement following treatment, are cumbersome and without an efficient solution procedure. In order to apply optimal control theory, Tsunokowa and Schofer optimised a continuous 'trend curve' that passed through the mid-point of each jump in the saw-tooth curve (see Figure 2.3). Using the same functions and parameters, Li and Madanat (2002) found that the 'trend curve' approximation of Tsunokowa and Schofer (1994) yielded sub-optimal policies. They argued that optimal control theory was unnecessary provided that the deterministic pavement deterioration and improvement models have the Markov property, that is, history does not influence system evolution. In other words, the deterioration rate and amount of improvement from a treatment at any point in time depends only on the pavement condition and decisions made at that point in time (or in the preceding period for discrete time). Despite the criticism from Li and Madanat (2002), Ul-Islam and Tsunokawa (2006) and Rashid and Tsunokawa (2012) again used the trend curve approach.

Li and Madanat (2002) found that, with an infinite time horizon, after the first rehabilitation, the system enters a steady state with a constant time interval between rehabilitations and constant roughness level at which the rehabilitations occur. With a finite time horizon, Ouyang and Madanat (2004 and 2006) found that the steady state was reached after just a few rehabilitations and stayed there until near to the end of the planning horizon. With identical pavement lifecycles, only the optimal time for a single cycle needs to be determined. These steady state models are similar to our simple model in Chapter 3, with a single treatment type and no change in user costs over time due to traffic growth. Abaza (2002) allowed for an integer number of rehabilitations between full reconstructions, effectively two treatment types, with each rehabilitation restoring the pavement to a lower condition than the previous one. Generally, treating time as discrete, rather than continuous, is much simpler and better suited to numerical solution (Belman 1961, p. 5).

Some models make treatment intensity a variable to be optimised. More intense treatments involve thicker overlays that improve road condition to greater degree but cost more. An upper limit is placed on overlay thickness, where it ceases to have any further effect on roughness. Treatment cost is made a linear function of overlay thickness. It seems that models of this type almost invariably find that the optimal strategy is a 'corner solution' rather than an intermediate standard that balances the marginal cost and marginal benefit of increasing overlay thickness. One corner solution is to resurface to the best achievable standard (Li and Madanat 2002, p. 531; Ouyang and Madanat 2006, p. 772). The other corner is to apply a very thin overlay in each time period. To prevent corner solutions with eight overlay thickness options, Tsunokawa and Ul-Islam (2003, p. 198), using the HDM-4 model, imposed a 5-year minimum overlay interval in the simulations "to avoid two consecutive condition-responsive overlays from being applied in an impractically short period of time".

Gu et al. (2012) showed that this conclusion does not apply where 'maintenance activities' such as crack sealing, are introduced into the model. Increased spending on 'maintenance activities' reduced the deterioration rate, which in turn postponed the economically optimal resurfacing time. The model then had to find an optimal trade-off between the intensities of two types of treatment, in this case, resurfacing and other 'maintenance activities'.

All the articles cited so far in this section on deterministic models with continuous pavement condition deal with a *single* segment in isolation without budget constraints. Sathaye and Madanat (2011) considered the case where rehabilitation treatments were optimised for *multiple* segments over an infinite time horizon

----------

[17] Another unusual feature of Friesz and Fernandez (1979), which is shared by Abelson and Flowerdew (1975), is inclusion of a demand relationship. Pavement condition affects traffic demand so that a better maintained pavement gives rise to a benefit from generated demand. This contrasts with the argument in Chapter 3, Section 3.2 that, over the relevant range of road conditions, effects on traffic demand are likely to be negligible and so can be omitted from models.

subject to a budget constraint for the annual maintenance costs summed over all segments. The decision problem was to minimise

$$\sum_{j} PVTTC = \sum_{j} \frac{(M_j + U_j)}{(1 - e^{-r\tau_j})} \qquad \text{subject to} \qquad \sum_{j} \frac{M_j}{\tau_j} \leq B$$

where the subscript *j* represents each segment, $PVTTC$ is the present value of total transport costs as defined previously in Chapter 3, *M* is rehabilitation cost, *U* is user cost between cycles, $\tau$ is the number of years between rehabilitations, and *B* is the annual budget assumed constant for all years.

The constrained minimisation problem can be expressed as minimising

$$L = \sum_{j} PVTTC + \lambda\left(\sum_{j} \frac{M_j}{\tau_j} - B\right)$$

where $\lambda$ is the Lagrange multiplier.

Assuming that the budget constraint is binding, "the optimal solution occurs where the marginal social benefit with respect to expenditure is equal for all *j*. … optimal values for $-\,dPVTTC_j/d(M_j/\tau_j)$ for all *j* will be equal to the Lagrange multiplier $\lambda$" (Sathaye and Madanat 2011, p. 1006). To solve a numerical example with three segments, they adjusted the trigger roughness level at which rehabilitation occurs for each segment so as to equate $-\,dPVTTC_j/d(M_j/\tau_j)$ for all segments. Sathaye and Madanat's approach is a precursor to the optimisation method in this report for minimising PVTTC subject to annual budget constraints set out theoretically in Chapter 3 and demonstrated with the case study in Chapter 6.

#### 4.3.1.2 Continuous pavement condition and probabilistic category

There is a large stochastic element in pavement deterioration caused by unpredictable and unmeasurable factors. Examples are the quality of the materials and workmanship in constructing and maintaining the pavement and drains, the characteristics of the sub-grade, the weather, and heavy vehicle loadings. There is also uncertainty about the effectiveness of maintenance treatments specified in models, that is, the amount by which they improve pavement condition measures (Qiao et al. 2018). If a deterministic deterioration algorithm is well calibrated, it should approximate the expected value of the probability distribution at each point in time. If probability distributions were assumed for one or more of the parameters that affect deterioration rates and treatment effectiveness, probability distributions for road condition measures and forecast spending needs could be derived. As with all forecasts of stochastic variables, the variance of the distribution of forecast outcomes would increase the further the system is projected into the future.

The list of probabilistic models with continuous pavement condition in Table 4.1 is short. One reason may be that, for a collection of road segments considered together, the expected outcome from a probabilistic model averages out to approximately the outcome from a deterministic model, so there is little to be gained by making the model probabilistic. Another reason is the analytical difficulty of managing such models using simulation methods due to the large number of possible outcomes to be modelled.

Li and Madanat (2002) undertook a simulation for a model with a single rehabilitation treatment type and simple deterioration function, $1.92e^{0.05t}$ m/km IRI (5.1% increase each year).[18] Varying the 0.05 coefficient uniformly over the interval 0.03 to 0.07, they did not find any significant change in the optimum (PVTTC-minimising) roughness level at which to rehabilitate the pavement. They concluded that their approach was "robust to the uncertainty in the deterioration process" (Li and Madanat 2002, p. 534).

Some articles solved specific problems involving probabilities that could not be addressed with a deterministic model. Gao and Zhang (2008) used 'robust optimisation' to identify options for the decision maker whereby higher agency costs could be accepted for a lower probability of violating a maximum roughness target. Ng et al. (2011) employed integer programming to find the set of treatment times and types that would minimise

----------

[18] Li and Madanat (2002) used the Quarter Index (QI) roughness measure. We have converted 25QI to 1.92m/km IRI using the relationship in Paterson (1986), 1 IRI = 13 QI.

undiscounted agency costs over the analysis period subject to a constraint that pavement condition would only fall below a minimum level with a specified probability. Their case study suggested that budget requirements can be more than twice the amount predicted by a deterministic model. With uncertainty about the effectiveness of each treatment and the amount of deterioration each year, it is necessary to implement more intense and more frequent treatments to meet the probabilistic pavement condition constraint.

These models highlight that fact that, seeking to minimise agency costs subject to minimum road condition constraints with a deterministic model will mean that, in practice, the condition of some segments of a network will fall below the constraint because of randomness in treatment effectiveness and pavement deterioration. Condition constraints may have to be set higher to allow for this, which will increase road agency costs. However, also in practice, decisions about treatments will be undertaken dynamically in response to new information becoming available. Being able to update the optimal strategy in response to new data from pavement inspections is a desirable feature of any pavement management system (Carnahan 1988, pp. 309-10). For a network of segments considered together, the additional costs of bringing forward treatments for segments in worse condition than forecast would be offset, at least in part, by cost savings from deferred treatments for segments in better condition than forecast.

## 4.3.2 Discrete pavement condition group — Table 4.2

Although the main measures of road condition (roughness, cracking, pavement strength and rutting) are continuous variables, in many optimisation models, road condition is treated as a discrete variable. For example, a pavement might be classed as either 'very poor', 'poor', 'fair', 'good', or 'very good' based on ranges of roughness and possibly in combination with one or more other condition measures, or visual inspection. Having a small number of discrete road condition states can greatly simplify optimisation and is essential for the Markov chain approach. Table 4.2 indicates that numbers of discrete states in the models in the literature surveyed ranges from 3 to 120. A state is defined as a combination of the specific levels of variables relevant to evaluating pavement performance (Golabi et al. 1982, p.10). A common discrete approach used with Markov models is to have ten states obtained by dividing the zero to 100 scale of the pavement condition index (PCI) into ten equal intervals.

According to Carnahan et al. (1987, p. 557), PCI determination can be subject to errors of up to ±5, so a more detailed state definition may be pointless. The models with higher numbers of states had states defined by different combinations of attributes such as bands of roughness and cracking.

### 4.1.1.1 Discrete pavement condition and deterministic category

Deterministic probabilistic models with discrete pavement condition are rare, with only three shown in Table 4.2. The reason may be that, with pavement condition taking on a limited number of discrete states, it is not analytically difficult to switch from a deterministic model to a probabilistic Markov model.

One of the earliest attempts at road maintenance optimisation for a network was made by Abelson and Flowerdew (1975) for roads in Jamaica. They relied on manual calculation of discounted costs over a 10-year analysis period to compare alternative maintenance strategies for roads in different condition states with different traffic levels to develop operating rules. The sum of road maintenance, vehicle operating and 'lost traffic' costs was minimised. It is unusual to include lost traffic, that is, traffic deterred by higher operating costs as road deteriorate. Lost traffic was valued at half the increase in vehicle operating costs consistent with the 'rule-of-half' used in cost–benefit analysis for consumers' surplus changes (ATAP 2022). Abelson and Flowerdew discussed how dynamic and integer programming might be applied to more complex models using computers. Commenting on their article, Gerchak and Waters (1978) suggested inclusion of traffic delay costs and a probabilistic Markov chain approach.

The modelling approach in de la Garza et al. (2011), while deterministic has all the characteristics of Markov approaches except that each pavement state deteriorates to a single pavement state rather than into multiple pavement states with different probabilities that sum to one.

**Table 4.2 Models with discrete pavement condition**

| Article | Objective function | Non-technical constraints | Optimisation method | Treatment types | Pavement condition | Analysis period (years) | Segments/families |
|---|---|---|---|---|---|---|---|
| **Deterministic** | | | | | | | |
| **Abelson & Flowerdew 1975** | min PVTTC | none | full enumeration to develop rules for various condition states | 7 | road classes | 10 | na |
| **Tack & Chou 2002** | max average PCI | annual budget | dynamic programming, and genetic algorithm | 3 | 10 states by PCI | 5 | 5, 10, 20, 40 tested |
| **de la Garza et al. 2011** | min number of lane miles in the 3 worst condition states less the weighted sum of lane miles in 2 best states, then budget | various | linear programming | 8 | 5 states | 15 | 1* |
| **Probabilistic (all with Markov chain approach)** | | | | | | | |
| **Golobi et al. 1982** | min PVAC | minimum performance standards | linear programming | 16 | 120 states **** | 5 | 9* |
| **Carnahan et al. 1987** | min undiscounted AC | minimum standard in final year with 95% probability | dynamic programming, | 5 | 8 states by PCI | 20 | 1* |
| **Feighan et al. 1988** | min expected PVAC | minimum performance standards | dynamic programming, | 3 | 7 states by PCI | 25 | 4* |

| Article | Objective function | Non-technical constraints | Optimisation method | Treatment types | Pavement condition | Analysis period (years) | Segments/families |
|---|---|---|---|---|---|---|---|
| **Feighan et al. 1989** | max average PCI units | none then annual budget | dynamic programming, then weighted effectiveness/cost ratio * | 3 | 10 states | 5 | 5* |
| **Grivas et al. 1993** | min PVAC | annual budget and overall network condition | linear programming | 7 but limited to 2 per condition state | 45 states | 5 and 10 | 6* = 2 pavement types x 3 traffic levels |
| **Butt et al. 1994** | max PCI units (area under curve) | none then annual budget | dynamic programming, then weighted BCR & IBCR * | 3 | 10 states | 5 | 13* |
| **Ravirala & Grivas 1994** | min wtg sum of deviations of objective functions from their respective goals *** | none | linear programming | 3 | 3 states | 1 | 4* |
| **Mbwana & Turnquist 1996** | min PVTTC | annual budget | linear programming | 4 | 10 states | na | 60 |
| **Li, Huot & Haas 1997b** | max cost effectiveness | annual budget | prioritisation | 4 | 10 states | 10 | 18 |
| **Davis. & Van Dine 1988** | min user costs (UC) | annual (max &/or min) or total budget, production capacity (max &/or min) | linear programming | 4 | 10 states | 5 | 12* |
| **Li, Hass & Huot 1998** | max effectiveness | annual budget, minimum performance standards | integer programming | 2 | PCI units | 5 | 5 |

| Article | Objective function | Non-technical constraints | Optimisation method | Treatment types | Pavement condition | Analysis period (years) | Segments/families |
|---|---|---|---|---|---|---|---|
| **Abaza & Ashur 1999** | max (min) prop of pavements in given good (bad) state / max wtd avg pavement condition / min undiscounted agency cost st pavement condition levels | annual budget | penalty method with pattern search | 3 | 5 states | 6 | 120* |
| **Durango & Madanat 2002** | min PVTTC | none | adaptive control * | 7 | 8 states by PCI | 25 | 1* |
| **Abaza et al. 2004** | max (min) prop of pavements in given good (bad) state / max wtd avg pavement condition / min undiscounted AC st pavement condition levels | annual budget | various non-linear techniques tested | 6 | 5 states | 6 | 1* |
| **Abaza & Murad 2007** | max pavement condition st budget constraints / min restoration cost st condition constraints | | linear programming | 8 | 6 states | 10 (2-yr intervals) | 1* |
| **Wang et al. 2007** | max average performance rating and min agency cost, undiscounted sum ** | min and max proportions of pavement in low and high condition states set to smooth changes over time | genetic algorithm | 5 | 45 states | 5 | 1* |

| Article | Objective function | Non-technical constraints | Optimisation method | Treatment types | Pavement condition | Analysis period (years) | Segments/families |
|---|---|---|---|---|---|---|---|
| **Wu & Flintsch 2009** | max wtd sum of percentages of pavement in various condition states (higher weights for better states) and minimise undiscounted sum of agency costs ** | min proportion of pavement in lowest state and weighted sum of proportions in each state; max proportion of network treated each year | not specified | 3 | 4 states | 10 | 1* |
| **Gao et al. 2012** | min average annual cost and max average proportion of network in very good condition state ** | annual budget and proportion of road network in first condition state | linear programming | 3 | 5 states | 10 | 1* |
| **Yeo et al. 2013** | min PVAC | minimum service levels and annual budget | dynamic programming, then pattern search * | 2 | 5 states | 40 | 20 to 2000 tested |
| **Medury & Madanat 2014** | min PVTTC | annual budget | Monte-Carlo simulation | 3 | 8 states | 15 | 10 to 1000 tested |

Notes: The time column in Table 4.1 has been omitted from Table 4.2 because all models are discrete.
* in the optimisation method column indicates a two-stage optimisation process. ** in the objective function column means bi-objective. *** multiple, goal programming, 35 goals
**** vector of four characteristics: cracking, change in cracking since previous year, roughness, index to the first crack
* in the segments/families column means multiple initial condition states for each segment or family
Abbreviations: max = maximise, min = minimise, PV = present value, st = subject to; TTC = total transport cost, UC = user cost, wtd = weighted, avg = average, prob = probability, PCI = pavement condition index

#### 4.1.1.2 Discrete pavement condition and probabilistic category

All the models in the literature with discrete pavement condition and probabilistic pavement deterioration use the Markov chain approach due to its ability to integrate pavement deterioration and treatment effects into a single transition probability matrix (de la Garza et al. 2011, p. 700). Golabi et al. (1982) was the first to implement this approach for road maintenance optimisation.

A stochastic process is considered a 'Markov process' if the probability of a future state in the process depends only on the current state and on actions taken in the immediately preceding period, not periods further in the past. A Markov chain is a series of transitions between states having the Markov property.

Central to Markovian pavement performance modelling is the specification of Transition Probability Matrixes (TPMs) that indicate the probability that a pavement in each state will change to another state. Transition probabilities are obtained from past data or expert judgement (Li Haas and Xie in TRR et al. 1997, p.71; Morcous and Lounis 2005, p. 131).[19] In the absence of any treatment, a pavement can only remain in the same state or deteriorate to a lower state. It can never rise to a higher state. Pavement condition is unlikely to decrease by more than 10 PCI units in a single year (Carnavan et al. 1987, p. 557). Hence, it is often assumed that a pavement cannot deteriorate by more than one state in a single time period (Ortiz-Garcia et al. 2006, p. 142). A maintenance treatment will cause a rise to a higher state. A different transition matrix is required for each treatment type including the null treatment.

Traditionally, TPMs are treated as being homogenous (stationary), that is, the road network will always deteriorate according to a fixed TPM. There is an implicit assumption that traffic and environmental conditions stay constant throughout the analysis period, which is not plausible for most real-world pavement situations (Li et al. 1996, p.204 and 1997b, p.9). Higher traffic loads and less favourable environmental conditions will increase the probabilities of deterioration. The problem can be addressed by using a non-homogenous (non-stationary) Markov process where the TPM changes over time.

The probabilities in a TPM can be interpreted as the proportion of pavements in a given initial state that will transition to each possible final state. Most of the case studies listed in Table 4.2 (starred in the 'segments/families' column) have the entire network or each family of segments with similar characteristics (traffic level, pavement type, deterioration rate) divided up by length between the various states at the start of the analysis period. The proportions of length in different states change each year. Two individual pieces of the road could be grouped together in the same state for one year and separated into the different states the following year, and vice versa.

The majority of Markov models estimate the fraction of pavement segments in a particular state to which a particular treatment is to be applied (Ng et al. 2011, p. 1327). In other words, the treatment recommendations need not be 'pure' strategies. A pure strategy would be: 'undertake treatment *a* when in state *i*'. The recommended treatment strategies are usually 'randomised', that is, 'when in state *a*, undertake treatment $a_1$ with probability $p_1$ and treatment $a_2$ with a probability $p_2$'. The probabilities can be interpreted as the proportions of lane-kilometres in state *i* for which treatments $a_1$ and $a_2$ are to be undertaken (Golabi et al. 1982, p. 15). Hence, the majority of Markov models cannot be used for planning maintenance on specific roads, such as segment *X* needs treatment *Y* in year *Z*. They are typically applied for estimating long-term budgets and making needs projections at the network level.

There are exceptions in the literature where a Markov model has been employed to determine optimal treatments for individual pavement segments, for example, Mbwana and Turnquist (1996). But even then, the model outputs are a set of recommended actions for each segment, one for each possible pavement condition state. Their approach was criticised by Wang et al. (2003, p. 2) for its complexity and assumptions.

----------

19 Khan et al. (2014) describe six ways to obtain TPMs from past data. An early example is Butt et al. (1987) who used constrained least squares.

### 4.3.3 Comparisons between discrete probabilistic and continuous deterministic models

There are differing views in the literature about the advantages and disadvantages of the two large categories in Tables 4.1 and 4.2, — deterministic models with continuous pavement condition and probabilistic models with discrete pavement condition.

Probabilistic models explicitly recognise the uncertainties in pavement deterioration. Carnavan et al. (1987, p. 556) argue that the Markov process seems superior to 'curve-fitting' approaches because it "introduces a rational structure for interpreting road condition data". Markov models start with current pavement condition data. It is not simple to fit a curve to current pavement condition when the actual current condition differs from what the model predicts given data inputs such as pavement age, initial pavement strength, cumulative standard axle loads and environmental conditions. An example would be data for a road segment indicating a relatively young pavement but with high roughness, inconsistent with the relationships in a deterministic model whereby only older pavements can have high roughness values.

Markov models work well with large-size networks when used to make projections for maintenance needs at network level (Morcous and Lounis 2005, p.131).

Ferreira et al. (2002b, p. 95) argued that probabilistic models lack solid theoretical foundations, in other words, they are purely empirical. Durango-Cohen (2007, pp. 494-5) criticised models that make condition a discrete variable on the grounds that "the partitioning process introduces forecasting errors and uncertainty". Both sources pointed out that Markov models have limited flexibility to add variables that describe pavement condition (for example, in addition to roughness, the type and extent of cracking, rut depth, profile, extent of surface patching). Adding variables to a Markov model causes computational effort to increase exponentially — the 'curse of dimensionality' discussed just below. For example, in Wang et al. (2007), three roughness levels (low, medium and high), three cracking levels and five index-to-first-crack measures gave rise to 45 condition states. Durango-Cohen (2007, pp. 494-5) further referred to empirical studies in pavement and bridge management that show that "physical deterioration of transportation infrastructure *may not be* Markovian" [italics in original]. According to Ferreira et al. (2002b, p. 95), these reasons possibly explain why road agencies often prefer pavement management systems based on deterministic pavement performance models rather than probabilistic models.

A number of studies have investigated the relationship between deterministic and probabilistic prediction models in pavement management. For example, Li et al. (1997a) developed a method to convert a deterministic model into a Markov model. Bekheet et al. (2008) compared the performance of a deterministic pavement prediction model and a Markov-based system. Validation was made of both systems against actual measured pavement condition data. The results showed that both systems performed well.

## 4.4 Objective function and constraints

### 4.4.1 Objective function

As discussed in Chapter 3, minimising the present value of total transport costs without budget or binding minimum standards constraints yields the most economically efficient solution. Of the 30 articles in Table 4.1 with continuous pavement condition, 15 minimised the present value of total transport costs. Only four of the 23 articles in Table 4.2 did this, showing it is relatively rare for models with discrete pavement condition. Seven articles of the 53 minimised the present value of agency costs and a further five did so without discounting. All those that minimised agency costs had minimum pavement condition constraints, which is the cost-effectiveness analysis approach discussed in Chapter 3. The full summary of objective functions in Tables 4.1 and 4.2 is shown in Table 4.3.

Many of the other articles listed optimised pavement performance or condition measures, reflecting the engineering frame of reference in the literature. Several maximised 'effectiveness' defined as the area under a performance curve (pavement condition as a function of time) multiplied by traffic level and segment length. This may or may not be time discounted. Multiplying by traffic level mirrors user costs being proportional to traffic numbers. One of the earliest maintenance optimisation articles, Mahoney et al. (1978) maximised gains of pavement rating points times survival probabilities. Morscous et al. (2005, pp. 132-3) discussed setting up the problem as maximising average network condition given annual budget constraints. They did not address how to differentiate between roads with different traffic levels. Davis and Van Dine (1988) was unusual in minimising user costs subject to road agency budget constraints.[20]

**Table 4.3 Summary of objective functions in literature survey**

| Table | Min PVTTC | Min PVAC | Min AC | Min AC and other | Other | Totals |
|---|---|---|---|---|---|---|
| **4.1** | 15 | 3 | 1 | 3 | 8 | 30 |
| **4.2** | 4 | 4 | 1 | 0 | 14 | 23 |
| **Totals** | 19 | 7 | 2 | 3 | 22 | 53 |

Note: Min = minimise, AC = undiscounted agency costs, Min AC and other = multiple objectives including undiscounted agency costs.

The lack of discounting in many objective functions also shows an absence of consideration of economic principles.

### 4.4.2 Constraints

Of the articles with continuous pavement condition in Table 4.1, 16 optimised without budget constraints but only four in Table 4.2 with discrete pavement condition. Budget constraints can be expressed in terms of a present value, an undiscounted sum over the analysis period, an average annual amount, or a series of annual amounts that can be uniform over time or varying. Annual budget constraints occur in seven articles in Table 4.1 and 12 articles in Table 4.2.

Condition constraints can be expressed as single pavement condition measures, average condition for a network, or the proportion(s) of a network meeting a condition. The latter occurs in Markov models, where a condition constraint can be expressed as a probability that a pavement segment will not fall below a given state. A constraint that an *x*% probability that the condition of any given segment will be above state *y* can be interpreted as *x*% of lane-kilometres must be kept in conditions above state *y* (Golabi et al. 1982, p. 10).

Availability of physical resources to undertake certain treatments (manpower, equipment and materials) may lead to further constraints (Chan et al. 2001). For example, Davis and Van Dine (1988) included in their maintenance optimisation model, minimum and maximum amounts of each treatment that could be deployed in each year. The minimums were “introduced to avoid a solution that calls for extreme shifts in pavement material production from year to year”.

Technical restrictions may be necessary to prevent unrealistic solutions. The purpose could be to prevent models extrapolating relationships beyond the range over which they apply. Road safety considerations might lead to setting upper limits on roughness, rut depth and skid resistance.

### 4.4.3 Multiple objectives

A number of later articles adopt multiple objective approaches. ‘Multi-objective programming’ identifies the Pareto frontier along which no objective can be advanced except at the expense of one or more of the others. Tables 4.1 and 4.2 contain eight examples of bi-objective studies. The most common bi-objective approach is to minimise agency costs together with either minimising user costs (for example, Meneses and Ferreira 2013) or maximising a road condition indicator (for example, Wu and Flintsch 2009). To reconcile *maximising* road condition with *minimising* agency cost, the problem can be written as minimise negative road condition plus minimise agency cost (for example, Gao et al. 2012). Wang et al. (2003) maximised treatment effectiveness

----------

[20] Fwa et al. (1998, p. 2) listed a number of other objectives for road agencies not already mentioned: maximisation of maintenance production, usage of allocated budget, usage of available manpower, and usage of available equipment; minimisation of manpower requirements, equipment requirements, and fluctuations in years demand for pavement expenditures.

(sum of improvements in pavement condition weighted by length, traffic and years life) and minimised the sum of disturbance costs to road users. An example with more than two objectives is Fwa et al. (2000) (not listed in Tables 4.1 or 4.2 because it deals with programming routine maintenance activities) with three objectives — minimise maintenance cost, maximise work production (days worked) and maximise network pavement condition.

The decision maker has to select the preferred point on the Pareto frontier. If some objectives are tightly constrained, that is, the decision maker has no interest in solutions that sacrifice an objective beyond a particular level, then some points on the frontier beyond the threshold level can be easily eliminated. However, in such cases, the objective could have been treated as a constraint. The multi-objective programming approach then has no advantage over optimising a single objective subject to constraints except to address the difficulties genetic algorithms (discussed below) have with identifying optimal solutions close to or on constraints.

Multi-objective programming may be advantageous if the constraints are 'soft' in the sense of being yet to be determined or open to negotiation. Model results can help decision makers and negotiators to understand the trade-offs between objectives. Any study that estimates optimal solutions for budget constraints set at a number of different levels is effectively locating points on the Pareto curve and could be considered bi-objective.

The most common optimisation method for multiple objectives is to maximise or minimise a weighted sum of the objective functions for each of the objectives. The weights do not have to sum to one as illustrated in Fwa et al. (1998). The optimisation model must be run a number of times with different weights, each run locating a new point on the Pareto frontier. Meneses and Ferreira (2013) used a genetic algorithm to identify the Pareto curve for the trade-off between undiscounted user costs and undiscounted agency costs summed over a 20-year analysis period. Points on the curve were found by minimising total weighted sums of user and agency costs attaching different pairs of weights that summed to one (that is, minimise $w \times$ user cost + $(1 - w) \times$ agency cost where $0 \leq w \leq 1$).

In the case of the goal programming approach of Ravirala and Grivas (1994 and 1995) and Ravirala et al. (1997), the aim was to minimise the weighted sum of deviations from the specified goals, for example, spending in each geographical division and percentages of mileage in each condition class in each division. This is not optimisation in the strict sense of the term as the goals are targets to be achieved rather than quantities to be maximised or minimised. The goals achievement method will locate a point on the Pareto frontier when the target point is set above the frontier. However, there is a risk that the target point may be set below the Pareto frontier and the model solves for an inefficient solution.

## 4.5 The curse of dimensionality

All attempts at maintenance optimisation at the network level face the 'curse of dimensionality' or 'combinatorial explosion' (Chan et al. 1994, p. 694; Pilson et al. 1999, p. 42; Abaza et al. 2001, p. 493).[21] Increasing the number of treatment types, segments in the network, and years in the analysis period increases the number of possible solutions exponentially.

Treating time in years as a discrete variable, the number of potential solutions for a single segment with $S$ treatment types plus the null treatment option over an analysis period of $T$ years is $(S + 1)^T$. This might be manageable for a single segment, but single segments cannot be optimised in isolation once annual budget constraints are introduced. Spending on one segment in one year reduces funds available for all other segments in that year.[22] Letting $N$ be the number of segments, the number of potential solutions is $(S + 1)^{T \times N}$ (Golroo and Tighe 2012, p. 328; Fwa et al. 1996, p. 246).

----------

21 The term 'curse of dimensionality' was introduced by Bellman (1961, pp. 8, 94 and 197.) when considering problems in dynamic programming. The number of samples needed to estimate an arbitrary function with a given level of accuracy grows exponentially with respect to the number of input variables (i.e. dimensionality) of the function.

22 For a present value budget constraint, the weighting method proposed in Chapter 3 and illustrated in the Chapter 5 case study enables individual segments to be considered in isolation, but this does not apply for annual budget constraints.

Tables 4.1 and 4.2 show that models that are high in one or more dimensions (number of treatment types, analysis years, and number of segments) compensate by being low in others. The most complex deterministic models with long analysis periods are often applied only to a single segment and may have just one rehabilitation treatment. The tables also show examples of models with shorter analysis periods, ten years and below, applied to over a hundred segments. Generally, Markov probabilistic models are applied with relatively short analysis periods.

There are ways to reduce dimensionality. The number time periods can be reduced by allowing maintenance actions only in specified years, for example, odd-numbered years 1, 3, 5, instead of every year. Only one instance of this was found of this in the literature review. Abaza and Murad (2007) assumed a 10-year analysis period divided into five time periods. In a few cases, a minimum time interval was set between periodic maintenance treatments (Abaza 2002; Tsunokawa and Ul-Islam 2003). Tack and Chou (2002 pp. 5-6) imposed a 4-year minimum allowable time constraint between major maintenance treatments ('frequency constraint'), as well as best and worst condition constraints for particular treatment types ('application constraints'). They noted that frequency and application constraints significantly reduced the number of treatment combinations to test. As mentioned above, treatment frequency constraints may also be imposed to prevent a corner solution where treatment intensity is variable.

The number of segments to be analysed can be reduced by aggregating segments into 'strategic sections' or 'families' with similar characteristics in terms of ranges of one or more of pavement condition measures, parameters in the deterioration algorithm (calibration and environmental coefficients), traffic level and vehicle mix.[23] Earlier models required high levels of aggregation, due to limited data and computing power. As data availability and computing technology have improved, there has been a trend toward greater disaggregation. Highly aggregated models can only produce maintenance policies — general guidance as to which treatments to undertake when and under which circumstances — not recommendations for specific individual road segments. Taking length-weighted averages of characteristics of small segments of road, when aggregating into larger segments, hides pieces of road in poor condition that may need treating sooner than predicted for the aggregated segment, leading to underestimation of future maintenance needs.

Having a short analysis period dramatically reduces the number of potential solutions but, as discussed in Chapter 3, due to treatments in the far future affecting the value of treatments in the near future, the length of the analysis period can affect results. Without a residual value or depreciation amount or a minimum condition constraint, the model has an incentive to allow a pavement to run down towards the end of the analysis period. It is not surprising then to find that the models with short analysis periods usually either have maximisation of some measure of pavement condition as their objective function or have minimum condition constraints.

The number of possible treatment types in the tables, not counting the null treatment option, ranges from one to 12. While the model realism can be enhanced by having a greater number of treatment types, it increases the likelihood that one or more of the menu of treatment types will never be selected by the model. This occurs where a treatment type cannot compete with (is dominated by) another alternative treatment type that offers a superior trade-off between effectiveness and cost. This was illustrated in Yepes et al. (2016) where no treatment types involving recycling techniques were included in the optimal solution. These treatment types were dominated by alternatives that had the same effectiveness (measured in years of service life increase) and a lower cost per square metre. The dominated treatments could have been excluded from the model with no effect on results.

Treatments can be considered as a menu of alternatives for which greater effectiveness in improving road condition and longevity can be purchased for greater cost. The choice between adjacent treatments in ascending order of cost and effectiveness can be very sensitive to their relative cost and effectiveness characteristics, so treatment choice becomes largely an artefact of the assumptions made about treatment

----------

[23] The term 'families' comes from the early road maintenance optimisation literature. According to Butt et al. (1987, p. 13), "A pavement family is a group of pavement sections with the same pavement type, the pavement use and the pavement rank." From Carnahan et al. (1987, p. 556), "… pavements are categorized so that roads with similar pavement construction methods, traffic loads and geographical location are collected into one class or 'family'…".

effectiveness and cost. Thus, increasing the number of treatment types in a model can have rapidly diminishing returns in terms of value added.

A way to dramatically reduce the number of possible solutions is to have the model choose between condition-responsive treatment rules instead of years in which particular treatments are implemented. An example of such a rule would be ‘rehabilitate as soon as roughness reaches 5 m/km IRI and resurface when cracking reaches 2% of surface area’. The model could test alternative trigger values for each segment, for example, rehabilitate at either 4 m/km, 5 m/km, or 6 m/km IRI. Triggers for different treatment types can be grouped together to form ‘maintenance policies’ (ARRB 2014). It is then no longer necessary to consider treatment implementation time as a variable to be optimised. Once a trigger is chosen, the implementation time for a treatment becomes endogenous to the model. Bhandari et al. (1987) did this to enable consideration of 30 policy alternatives for paved roads.

Use of triggers will be less satisfactory where traffic is growing. Traffic growth alters economically-optimal condition triggers because higher maintenance standards tend to be economically warranted for more highly trafficked roads. In other words, a trigger level set for a segment over the entire analysis period cannot allow for the changing economic value of treatments over time as traffic levels grow.

## 4.6 Optimisation methods

### 4.6.1 Survey of methods

The studies listed in Tables 4.1 and 4.2 feature a wide variety of optimisation methods. The earliest attempt at road maintenance optimisation by Abelson and Flowerdew (1975) used full enumeration to develop rules for various condition states for road maintenance in Jamaica over a 10-year analysis period. Since then, many techniques have been employed to solve pavement optimisation problems. They include

- Full enumeration of all possible solutions
- Mathematical programming methods, such as
    - optimal control theory
    - gradient method
    - linear programming
    - non-linear (including convex) programming
    - integer programming
    - dynamic programming
- Heuristic methods, such as
    - genetic algorithms
    - pattern search
    - greedy algorithms
    - simulated annealing
    - Monte-Carlo simulation.

Full enumeration involves computing the net present values of all feasible solutions. It guarantees finding the optimal solution (Odoki and Kerali 2006, G1-23). Whether or not it is feasible depends on the number of solutions to be evaluated and computation speed.

Descriptions of the various mathematical programming techniques are widely available and so are not given here. Table 4.1 shows that models with continuous pavement condition and continuous time are restricted in the types of optimisation approaches available — mainly optimal control and differentiation. For discrete time, Tables 4.1 and 4.2 show that linear, integer and dynamic programming are common approaches. In the cases of integer and linear programming, whether or not a treatment of a given type is implemented in a particular year on a particular road segment can be represented in the objective function as a variable that can take on only values of zero or one.

Generally, while mathematical programming methods provide optimal solutions, they are unable to deal with large networks, where the large number of decision variables render computation time impractically long (Torres-Machí et al. 2014, p. 5). Ouyang and Madanat (2004, p. 347) observed that mixed integer non-linear programming incurs prohibitive computational costs when the problem scale increases. Tack and Chou (2002, p. 8) and Medury and Madanat (2013, p. 137) remarked that the backward recursive dynamic programming approach guarantees finding the optimal policy, but suffers the 'curse of dimensionality' in that the computational costs increase exponentially with the number of possible states and decision options.

The alternative is heuristic search methods. They can provide satisfactory solutions in a reasonable amount of time for optimisation problems where full enumeration and mathematical programming approaches are impractical or would take too long. For example, Ferreira et al. (2002a) was able to solve a 27-segment problem with a four-year analysis period using mixed integer programming but had to rely on a heuristic technique for a 254-segment problem.

### 4.6.2 Genetic algorithms

Tables 4.1 and 4.2 show that genetic algorithms (GAs) tend to be the most common heuristic method employed for road maintenance optimisation. They are described here in some detail because one is employed in the case study in Chapter 6. They also illustrate characteristics of the broad heuristic family of algorithms.

GAs, first introduced by Holland (1975) and further elaborated by Goldberg (1989), are based on Darwinian evolutionary principles. Since the early 1990s, various GA methodologies have been developed to solve increasingly complex optimisation problems. The first application to road maintenance optimisation was reported in Chan et al. (1994) and Fwa et al. (1994a, 1994b).

GAs work on a coding of the parameters, not the parameters themselves (Fwa et al. 1994a, p. 31). They commence by generating a randomly selected parent pool of feasible solutions. The parameters describing each solution are encoded into a genetic representation or 'chromosome', which is an array of bits of information. Parent chromosomes are selected according to their 'fitness', usually measured by the objective function. The fitter chromosomes are mated to produce offspring, generating a second generation of chromosomes. Thus, through an iterative process, where the fitter individuals are allowed to mate more often, successive generations grow in fitness. Selection for mating typically involves a random element with the fitter individuals having a higher probability of selection. Mating, called 'crossover' or 'recombination', involves two parents swapping bits to produce two offspring. There are a variety of different ways to choose which bits to swap, the simplest being all bits to the right of a randomly selected crossover point.

Along the way, mutations are made to gene pool members to help move out of local optimums. Randomly selected bits are altered, for example, by swapping bits between positions or by replacing the value of one bit with a new randomly-selected value. Parameters such as the mutation and crossover probabilities and the population size can be adjusted to suit the particular problem. A mutation rate that is too low can lead to failure to find good solutions further away from those represented in the population pool, while too high a mutation rate can lead to loss of good solutions. A crossover rate that is too high can lead to premature convergence of the GA to a particular solution, ruling out the possibility of finding alternative good solutions.

The process continues until a termination condition is reached and the best solution out of all generations is chosen. Termination could be set to occur after a given amount of computer run time or a given number of generations or a criterion met for the best solution found. Successive trials tend to show diminishing returns in the improvements to the best solution out of each generation, so a plot of the best solution from each generation tends to plateau.

GAs differ from traditional optimisation techniques in a number of ways. GAs retain in memory at any one time a pool of feasible solutions rather than one single solution. The search process is not gradient-based so there is no requirement for differentiability or convexity of the objective function (Fwa et al. 2000). The solution found is not guaranteed to be either a local optimum or the global optimum. How close it is to a local optimum depends on the termination condition, and there remains the possibility of other local optimums that have not been explored. Fwa et al. (1998, p. 4) argued that there exist many near-optimal solutions in a pavement management programming problem and these near-optimal solutions are practically as good as the optimal solution.

GAs have been found to work well for unconstrained problems where all possible solutions are considered feasible. However, road maintenance optimisation problems invariably have technical constraints and usually resource constraints. For GAs, the most common way to handle constraints is the ‘penalty method’, whereby a penalty is imposed on an infeasible solution that reduces its fitness. The penalty is usually a function of the extent to which the constraint is violated, for example, the more a solution is over budget, the higher the penalty. There are no rules for the form and parameters of the penalty function, but the penalties have to be large enough to prevent the population becoming dominated by infeasible solutions but not so large as to trap a search in a local optimum, where a single valid but poor solution dominates the population. The advantage of a penalty that only disadvantages a constraint-violating solution over the ‘death penalty’ that eliminates it from the population pool altogether, is that the pool retains the information contained in the constraint-violating solutions. (Chan et al. 1994, p. 702; Pilson et al. 1999, p. 43; Chan et al. 2001, pp. 180-3; Yeniay 2005) Other ways to handle constraints are decoder or repair algorithms that avoid creating invalid offspring (Fwa et al. 1994a, p. 35).

Because GAs retain multiple solutions in memory, they are well suited to solving multi-objective problems (Konak et al. 2006). As an alternative to the weighting method discussed above for multi-objective optimisation, genetic operators can be modified to evolve a set of non-dominated solutions in a single run (Deb 1999). Deb (1999) and Konak et al. (2006) discuss some of the problems and design issues with multi-objective GAs.

A drawback of heuristic methods is that the analyst does not know how close the best solution is to the optimal solution. Ideally, results would be periodically compared with results from a mathematical programming method to ensure the solutions are optimal or near-optimal (Zimmerman 1995, p. 11). This is practical where the number of potential solutions is not too large. Chen et al. (2015) compared a GA in bi-objective optimisation with their own ‘dichotomic’ approach and found that that the GA could only identify seven solutions on the Pareto frontier when 50 segments were analysed, and one when 400 or more were analysed. They concluded that the GA solutions become worse as the number of segments increases.

### 4.6.3 Two-stage approaches

Tables 4.1 and 4.2 feature six articles (starred in the optimisation method column) that employ two-stage approaches to managing the problem of dimensionality when optimising subject to annual budget constraints.[24] Torres-Machí et al. (2014) termed these ‘sequential’ or ‘iterative’ approaches, in contrast to ‘holistic’ approaches that rely on a single round of optimisation. Medury and Madanat (2014) used the terms ‘top-down approach’ for a single stage optimisation and ‘two-stage bottom-up approach’.

In the first stage, each segment is considered in isolation in the absence of budget constraints. Under the ‘sequential’ approach of Torres-Machí et al. (2014), the optimum solution for each segment in the absence of budget constraints was identified in the first stage. Budget constraints were introduced in the second stage. Segments requiring maintenance treatments were ranked using a prioritisation method enabling lower-ranked segments to be progressively excluded up to the point where budget constraints were met.

Under the ‘iterative’ approach of Torres-Machí et al. (2014) and the ‘two-stage bottom-up approach’ of Medury and Madanat (2014), in addition to the best solution for each segment, the second, third and so on best solutions were identified for each segment. It was then possible in the second stage to select a second or third-best solution for one segment over the best solution for another segment.

Two-stage approaches can dramatically reduce the number of combinations to be considered. As previously explained, with *S* treatment types plus the null treatment option over *T* years and *N* segments, there are $(S + 1)^{T \times N}$ possible solutions. Two-stage approaches have $N(S + 1)^T$ solutions to assess in the first stage, because there are $(S+1)^T$ solutions for each one of *N* pavement segments (Torres-Machí et al. 2014). Three of the six articles identified in Tables 4.1 and 4.2 as using two-stage approaches employed dynamic

----------

[24] In classifying of articles as one- and two-stage approaches in Tables 4.1 and 4.2 , multiple objective approaches in which the second stage is for the decision maker to choose between alternative points on a Pareto frontier, were classed as one-stage.

programming in the first stage, which is feasible for a single segment considered by itself. The others used heuristic approaches for the first stage.

As discussed above, prioritisation methods are unable to find economic optimums over long time spans because they consider only the budgetary impact of the earliest treatment of an option consisting of a series of treatments in various future years. There is further loss of optimality where the sequential approach passes only a single highest priority solution for each segment from the first stage to the second stage. For some segments, there may be inferior solutions not retained that are better than the highest priority solutions for some other segments. The iterative approach addresses this limitation but still relies on prioritisation. In their numerical example, Medury and Madanat (2014) showed that as the budget constraint becomes more severe, the single stage approach outperforms the two-stage approach.

The two-stage approaches in the literature to date offer ways to manage the curse of dimensionality and find solutions that are better than many other solutions. However, the solutions are not optimal and it is not known how far away they are from the optimum.

## 4.7 Conclusion

With almost all the literature on road maintenance optimisation coming from the civil engineering discipline, a minority of articles take the pure economic approach of minimising the present value of total transport costs. A small number take the cost–effectiveness analysis approach of minimising agency costs subject to road condition constraints. Annual budget constraints are common in the literature.

Most models in the literature are either deterministic, with pavement condition as a continuous variable, or probabilistic, with pavement condition as a discrete variable. The approach in the present report is deterministic, with continuous pavement condition. The theoretical discussion in Chapter 3 and the optimisation methodology in Chapter 6, for optimisation subject to budget constraints, applies the method of Lagrange multipliers, which has only been used once before in the road maintenance literature, by Sathaye and Madanat (2011). However, it is practically the same as the penalty method used for treating constraints when applying genetic algorithms to maintenance optimisation problems. The Lagrange method is well-known to economists because of the central role it plays in the mathematical formulation of the micro-economic theory of consumer and producer behaviour. The present report shows the link between the Lagrange multiplier or penalty factor and the marginal benefit–cost ratio.

A few articles in the literature apply two-stage optimisation approaches to address the curse of dimensionality. In the first stage, optimal solutions are identified for individual segments in the absence of budget constraints. Budget constraints are introduced in the second stage when a prioritisation method is applied. Prioritisation methods are not guaranteed to find optimal solutions. The case study in Chapter 6 applies a multi-stage approach that does not rely on prioritisation.

# 5. Case study without annual budget constraints


## Summary

A case study was undertaken using a database of 2034 segments of road with a total length of 1977 kilometres drawn from the non-urban parts of seven different highways supplied by an Australian state government road agency.

BITRE engaged the Australian Road Research Board (ARRB) to curate the data into a form suitable for maintenance modelling, provide HDM-4 calibration information, and to estimate economically optimal spending (minimising the present value of total transport costs (PVTTC)) using HDM-4 and their own modelling approaches.

BITRE developed its own model based on simplified HDM-4 pavement deterioration algorithms. The optimisation approach was full enumeration of all possible solutions subject to a minimum time interval between treatments. With three periodic maintenance treatment types, a minimum of eight years between treatments and a 40-year analysis period, there were up to 581,485 solution options for each segment. For the first 20 years, economically optimal spending for periodic (excluding routine) maintenance for the whole network was estimated at $1505 million, an average annual amount of $75 million. By length, 37% of the network would be rehabilitated over the 20 years. First-year optimal spending was estimated at $186 million.

The case study data was then used to illustrate the cost–effectiveness approach of minimising the present value of road agency costs (PVAC) subject to minimum standard constraints in the form of maximum permitted roughness levels. The maximum roughness level constraint declined with traffic level and the number of heavy vehicles, patterning the relationship between traffic and economically optimal road standard. Compared with the economically optimal PVTTC-minimising solution, 20-year spending was lower by 18% achieved by halving the amount of the rehabilitation work by road length in the first 10 years. Roughness was, on average, higher, which imposes additional costs on road users. Each dollar of PVAC saved, on average, cost users $2.30 compared with the PVTTC-minimising economic optimum.

The unconstrained PVTTC-minimising result implies a marginal benefit–cost ratio (MBCR) of one. Results were estimated by raising the MBCR in steps from 1.5 to 25. Increasing the MBCR pushes road agency maintenance spending into the future as well as reducing it in total. Setting a target MBCR above one is equivalent to minimising PVTTC subject to PVAC constraints except that the penalty factor or Lagrange multiplier is set exogenously, and the PVAC constraint is an output of the model rather than an input. The upward adjustment to the MBCR to save each additional increment of PVAC is at first gentle as the model initially delays non-urgent treatments. But the rise soon becomes extremely steep as the model is forced to delay increasingly-needed maintenance treatments.

A number of sensitivity tests were undertaken. Raising the discount rate reduced agency spending pushing maintenance activities into the future. Failing to include safety in user costs reduced and delayed recommended maintenance spending. If computer run times for the model need to be reduced, the sensitivity tests showed that it is better to retain the longer 40-year analysis period and skip testing options with treatments in some years in the later part of the period than to shorten the analysis period to 30 or 20 years.


## 5.1 Introduction

Most of the road maintenance optimisation articles reviewed in Chapter 4 feature a case study to demonstrate and test their methodologies. In the present report, the case study brings together and illustrates many of the concepts discussed in the previous chapters including

- data requirements and components maintenance optimisation models from Chapter 2
- the conceptual framework in Chapter 3, including modelling issues such as treatment of budget constraints and depreciation at the end of the analysis period, and

- optimisation discussed in Chapter 4.

The case study provided an opportunity to develop and test a multi-stage optimisation method that can handle a very large database of road segments.

The case study is divided into two parts. This chapter (Chapter 5) introduces the case study and presents results for minimising PVTTC and PVAC without budget constraints followed by minimising PVTTC with present value budget constraints. The more difficult modelling task of optimising subject to annual budget constraints follows in Chapter 6.

An Australian state government road agency supplied a database of road traffic and condition information. BITRE engaged ARRB to curate the data into a form suitable for maintenance modelling including adding climate data, to provide technical advice, and to model the data themselves (Toole and Roper 2014). ARRB's results for minimising PVTTC without budget constraints are presented in Appendix D and compared with BITRE results. The pavement deterioration and user cost algorithms in the case study model have already been presented in Chapter 2. Further details of the model are set out in this chapter, including the assumptions about maintenance treatments.

Results are presented in detail for the central scenario of minimising PVTTC without budget constraints in a range of ways to show their implications for maintenance activities, spending needs and the condition of the road network for the following 20 years. Results are then presented for minimising PVAC subject to maximum roughness constraints.

The method of minimising PVTTC subject to minimum present value budget constraints developed in Chapter 3 is illustrated for a range of target MBCRs.

Sensitivity tests are reported, including tests of ways to reduce computer run times.

## 5.2 Case study data

The case study data consisted of 2034 road segments with a total length of 1977 kilometres of the non-urban parts of seven different highways supplied an Australian state government road agency. For undivided roads, each segment represented the entire width of the road with lanes in both directions. For divided roads (lanes in the two directions separated by a median strip), each direction of the carriageway was represented as a separate segment. Individual segments ranged in length from 0.015 to 17.7 kilometres, with an average length of 0.972 kilometres. The segments at the low end of the length range were unrealistically short for maintenance treatment purposes because it would be uneconomic to apply a major treatment to a short length of road in isolation, but short segments were not combined with adjacent segments in order to demonstrate modelling with a large database.

Table 5.1 summarises the traffic and condition measures for the database with length-weighted averages and indicators of the range of variation. Average annual daily traffic (AADT) levels ranged from 1,000 to 17,500 vehicles per day. Some segments were in extremely poor condition with the worst road condition measures in the database being cracking at 94%, pavement strength at 2.8 adjusted structural number, and roughness at 7.6 m/km IRI. The minimum pavement age of 9 years indicates that no segments had been rehabilitated in almost a decade.

**Table 5.1 Case study network characteristics for all segments in year zero**

| | Minimum | 50th percentile | Length weighted average | Maximum |
|---|---|---|---|---|
| **AADT** | 1018 | 2829 | 3468 | 17,538 |
| **% Heavy vehicles** | 10% | 21% | 24% | 37% |
| **Annual light vehicle growth rate (%)** | 1.4% | 1.8% | 1.9% | 2.4% |
| **Annual heavy vehicle growth rate (%)** | 1.9% | 2.3% | 2.3% | 2.6% |
| **Surface age (years)** | 0 | 15 | 13 | 29 |
| **% Cracked** | 0 | 3% | 7% | 94% |
| **Pavement age (years)** | 9 | 25 | 28 | 37 |
| **Design pavement strength (SNP)** | 5.3 | 6.0 | 6.1 | 7.3 |
| **Pavement strength (SNP)** | 2.8 | 5.1 | 5.2 | 6.9 |
| **Rut depth (mm)** | 1.0 | 6.0 | 6.3 | 19.0 |
| **Roughness (m/km IRI)** | 1.0 | 2.0 | 2.1 | 7.6 |

Notes: Length weighted average = $\sum w_i x_i / \sum w_i$ where $w_i$ is the length of segment *i*.
50th percentile = 50% of the entire road length in the database was above, and 50% below, the amount shown.
AADT = average annual daily traffic, SNP = adjusted structural number, IRI = international roughness index.

The database lacked data on design pavement strength and measured pavement strength, information that is essential for modelling deterioration. Values for design pavement strength were generated by assuming that segments were originally constructed or previously rehabilitated to Austroads design standards. Design pavement strength was estimated from 'design traffic', assumed to be the number of millions of equivalent standard axles forecast for the 30 years following construction or rehabilitation (*MESA*). The design adjusted structural number ($SNP_0$) was given by:

$$SNP_0 = 4.7\ MESA^{0.1033} \tag{5.1}$$

For the purpose of estimating design pavement strengths, traffic levels were projected backward to the year of construction given by the pavement age. To allow for pavement strength to decline over the period between construction or the last rehabilitation and the start of the analysis, equation 2.2 in Chapter 2 was employed.

Of the 2034 segments, 493 were asphalt mix pavements and 1541 were surface treatment (sprayed seal) pavements, using the two-way classification for bituminous surfaces in HDM-4 (Odoki and Kerali 2006 Part C). By length, 164 km (8.3%) were asphalt mix pavements and 1812 km (91.7%) were surface treatment pavements.

## 5.3 BITRE modelling

To gain the flexibility to test a range of different approaches to maintenance modelling, BITRE developed its own model. The first version was implemented in an Excel workbook with a Visual Basic macro managing the progression through the database, one segment at a time, testing all the options, and recording the results. To address impractically long run times, the model was recoded in Mathematica, with checking to ensure both versions produced identical results.

Compared with models in the literature, the BITRE case study modelling

- was deterministic with discrete time intervals of one year
- minimised PVTTC subject to loose minimum pavement condition constraints imposed to ensure technical realism, or alternatively minimised PVAC subject to pavement roughness constraints set to take into account the interests of users

- treated pavement condition measures as continuous variables.

### 5.3.1 Pavement deterioration

To model pavement deterioration, HDM-4 deterioration algorithms (Morosuik 2004; Odoki 2006) were employed, with the addition of ARRB's relationship for pavement strength as a function of age given in equation 2.2 in Chapter 2. The deterioration algorithm was similar to Paterson's (1987) incremental model in equation 2.1 in Chapter 2 whereby the roughness increase in a single year is the sum of roughness increases from four processes — cracking, pavement strength decline, potholing and rutting — plus an environmental component dependent on climate. The main processes and inputs were illustrated in Figure 2.1 in Chapter 2.

Cracking was assumed to start (rising above 0.5%) after the surface age reached 12 years for surface treatment pavements and 16 years for asphalt mix pavements, by which time the bitumen has begun to oxidise. Cracking then proceeded along an S-curve from zero to 100%. Cracking has direct impacts on all the other processes. Potholing only commenced at relatively high levels of cracking and potholes were assumed to be quickly patched. Pavement strength declined with age and the rate of decline was accentuated for cracked pavements and wetter climates. Rutting was affected by cracking, pavement strength, axle loadings and climate. From mean rut depth, the standard deviation of rut depth was estimated, which contributes to roughness.

### 5.3.2 Treatments

BITRE's model dealt with periodic maintenance only. Routine maintenance could be estimated as fixed annual amounts per square metre of pavement or kilometre of road length outside the model. There were three types of treatments, which are shown in Table 5.2 in ascending order of cost per square metre and effectiveness in improving road condition. Treatment costs were higher for asphalt mix pavements relative to surface treatment pavements. Costs of resurfacing and shape correction increased linearly with the level of cracking. Rehabilitation costs rose with the design standard and there was a cost penalty if roughness rose above 4.1 m/km IRI at the time of rehabilitation because an additional layer of pavement needed to be replaced. The cost of treatments on surface treatment pavements was assumed to transition to the cost for asphalt mix pavements above a 30-year design traffic level of 20 million equivalent standard axles (MESA) or a design strength of 6.4 for the adjusted structural number.

Three treatment types may appear few but the literature survey summary Table 4.1 showed it to be mid-range for deterministic models with continuous pavement condition in the literature. The drawbacks of having a high number of alternative treatment types were discussed in Section 4.5.[25]

The following technical restrictions were imposed in the model.

- Resurfacing treatments were not permitted for levels of cracking above 25%.
- Resurfacing with shape correction treatments were not permitted for levels of cracking above 40%.
- Roughness was not permitted to rise above 6.3 m/km IRI.
- Pavement strength was not permitted to fall below an adjusted structural number of 3.0.

ARRB's advice was that roughness should not be permitted to exceed 5.2 m/km, that adjusted structural number not fall below the maximum of 3.0 and 60% of the design value, and that cracking not exceed 50%. There were a significant number of segments in the database already beyond or close to these limits and hence would require rehabilitation in the first year of the analysis period if these restrictions were implemented. As an aim of the modelling exercise was to explore costs of deferring maintenance, the restrictions recommended by ARRB were relaxed to enable greater flexibility to defer maintenance. It was

----------

25 The drawbacks of a greater number of treatment types are: (1) it increases the number of options to test by an exponential amount, (2) high likelihood of some treatments never being selected because they are dominated by others, and (3) high sensitivity of the choice of treatments to cost and effectiveness assumptions.

found that once the model had treated the segments that were initially outside the technical restrictions, economic optimisation in subsequent years made the restrictions redundant.

**Table 5.2 Treatment and cost assumptions**

| Treatment | Description | Computer code to estimate cost ($ per square metre) |
|---|---|---|
| **Resurface** | ST: 10mm overlay;<br>AM: 20mm overlay; | 0.3 * ACA + If [PavementType == "AM", 26.5, 4.5] |
| **Resurface with shape correction** | ST: 20mm overlay;<br>AM: 40mm overlay | 0.3 * ACA + If [PavementType == "AM", 43.5, 26.5] |
| **Rehabilitation** | Replace layers as required to raise pavement strength to design level of: Max [5.5, 4.7 * MESA^0.1033]. | If [SNP > 6.4, 56.3 * SNP – 206.7,<br>If [PavementType == "AM", 76.8 * SNP – 338,<br>115.6 * SNP – 586.5]]<br>+<br>If [IRI > 4.1,<br>If [SNP > 6.4, 9.76 * ( IRI – 4.1),<br>(2.73 * SNP – 7.72) * ( IRI – 4.1)],<br>0] |

Note: Values for ACA, MESA, SNP and IRI are all as at the time just prior to treatment.
ST = surface treatment, AM = asphalt mix, ACA = per cent cracked, MESA = millions of equivalent standard axles forecast for the next 30 years, SNP = adjusted structural number, IRI = m/km roughness.

Each treatment had 'reset' impacts, which are summarised in Table 5.3. All treatments reset surface age and cracking to zero. Rehabilitation reset roughness to 1.2 m/km IRI, the level of a new pavement, and raised pavement strength to the design level. Resurfacing with shape correction reduced roughness by a significant amount, typically by around one m/km IRI.

**Table 5.3 Summary of treatment reset impacts**

| | Surface age | Cracking | Pavement age | SNP | Rutting | Roughness |
|---|---|---|---|---|---|---|
| **Resurface** | 0 years | 0% | no change | add 0.1 | no change | no change |
| **Resurface with shape correction** | 0 years | 0% | no change | add 0.2 | subtract 1.5 mm | HDM-4 formula; approximately subtraction of 1 m/km IRI |
| **Rehabilitation** | 0 years | 0% | 0 years | required design level | 1.0 mm | 1.2 m/km IRI |

## 5.3.3 User costs

The relationship between roughness and road user costs in the BITRE model was discussed in Chapter 2 and illustrated in Figure 2.4.

## 5.3.4 Other assumptions

The analysis period was set at 40 years and the discount rate at 4%. Treatments in the model were assumed to be implemented at the end of each year.

The role of residual value or depreciation as an approximation of the PVTTC for the years after the end of the analysis period to infinity was discussed in Chapter 3. The 40-year analysis period pushes the end of the analysis period far enough into the future to limit the effects of the finite analysis period on the first 10 or

20 years, which are of most interest. In the BITRE model, depreciation was added to road agency costs at the end of the analysis period. Depreciation for each segment at the end of the analysis period was calculated based on surface age and roughness together with treatment costs. The maximum depreciation was set at the cost of rehabilitating a pavement with a roughness of 5.2 m/km and with a design pavement strength set for the traffic at the end of the analysis period. The approach to calculating depreciation in the model is described in detail in Appendix C.

### 5.3.5 Data requirements

Data fields, including HDM-4 calibration coefficients starting with '*K*', required by the model were

- *Traffic (average annual daily traffic ( AADT))*: car, light/med rigid, heavy rigid, articulated truck, combination truck
- *Traffic growth rates (linear projection)*: light vehicles, heavy vehicles
- *Climate / environment*: *m* (environment coefficient), $K_{gm}$, Thornthwaite moisture index (TMI), mean monthly precipitation (MMP)
- *Road characteristics*: length (km), carriageway width (m), paved area ($m^2$), pavement type (AM/ST), divided/undivided/freeway, and
- *Road condition*: surface age, cracking (%), $K_{cpa}$, pavement age, design and current pavement strengths (adjusted structural number), $K_{gs}$, rut depth (mm), $K_{rst}$, roughness (IRI).

The HDM-4 model requires significant configuration and adaptation to conditions in the country or region for which it is to be used. HDM-4 has calibration coefficients that modify the rates of progression of cracking, rutting and roughness and the time to initiation of cracking. Separate sets of calibration coefficients were supplied by ARRB for seven zones with differing road designs, land types and climates. For each of the seven zones, there were also specified values for the Thornthwaite moisture index, the HDM-4 environmental coefficient '*m*', and mean monthly precipitation.[26]

## 5.4 Optimisation without budget constraints

### 5.4.1 Representation of treatment 'options'

In the model's computer program, a treatment 'option' was represented as a two-part nested list. The first part of the list was a list of treatment times from years one to 40. The second part was a list of corresponding treatment types, coded 1 for a resurface, 2 for a resurface with shape correction and 3 for a rehabilitation. To illustrate, the option {{5, 18, 35}, {2, 1, 3}} implies a resurface with shape correction (treatment type 2) will be undertaken in year five, a resurface (treatment type 1) in year 18, and rehabilitation (treatment type 3) in year 35. The list {{3, 11, 20, 30, 39}, {1, 1, 3, 1, 2}} implies five treatments will be undertaken over the analysis period with resurfacings in years 3, 11 and 30, a rehabilitation in year 20 and a resurface with shape correction in year 39. The list for no treatments whatsoever is {{ }, { }}. Both parts of the list must be of equal length.

### 5.4.2 Optimisation method

Without annual budget constraints, each individual segment can be processed in isolation because spending on one segment in any year has no impact on funds available for the other segments in that year. But this is

----------

26 The table here shows the magnitudes of the HDM-4 calibration coefficients and climate data items.

| | $K_{cpa}$ AM | $K_{cpa}$ ST | $K_{gs}$ | $K_{gm}$ | $K_{rst}$ | m | MMP | TMI |
|---|---|---|---|---|---|---|---|---|
| **Minimum** | 0.7 | 0.26 | 0.3 | 0.3 | 1.25 | 0.03 | 32.14 | –17.18 |
| **Length weighted average** | 0.7 | 0.29 | 0.66 | 0.65 | 1.76 | 0.03 | 45.91 | 4.60 |
| **Maximum** | 0.7 | 0.35 | 1.19 | 1.19 | 2.2 | 0.04 | 71.03 | 41.45 |

AM = asphalt mix pavement; ST = surface treatment pavement; MMP = mean monthly precipitation; TMI = Thornthwaite moisture index not sufficient to address the 'curse of dimensionality'. For a single segment, with four treatment types, including the null treatment, and 40 years, there are $4^{40} \approx 10^{24}$ options.[27]

In early attempts to find optimum solutions, it was found that, faced with the huge number of possible solutions and with a starting point far away from the optimum, the genetic algorithm software was unable to find a solution that could not be significantly improved by making manual changes.

Ways to reduce the number of solutions to a manageable level were discussed in Section 4.5. One of these is to set a minimum time interval between treatments. Specifying a minimum time interval between treatments is a reasonable restriction to impose on solutions because the economically optimal solution is unlikely to involve one periodic maintenance treatment implemented within a few years or less of another unless the treatments have very low implementation costs.

The first treatment implemented during the analysis period can occur in any year from 1 to 40, or no treatment may in implemented in any year. After the first treatment has occurred, a minimum time interval of eight years applied. For example, a solution option with treatment times {10, 16, 30} would not be tested, while {10, 18, 30} would be. The 8-year minimum time interval is well under the 12-year period before crack initiation assumed for surface treatment pavements and the 16-year period for asphalt mix pavements. The number of treatments in each option then varies from zero to five. Five treatments, eight years apart, is the maximum that can fit within the 40-year analysis period, for example, treatments in years {1, 9, 17, 25, 33} or {8, 16, 24, 32, 40}. With a minimum interval of eight years between treatment times, there are 7837 possible combinations of treatment times that will fit within a 40-year analysis period.

For each combination with *n* treatments over the analysis period from zero to five, there are $3^n$ permutations of treatment types giving rise to 581,485 possible options to test for each road segment. With this number of options, the full enumeration approach is feasible for optimisation as long as each segment can be considered in isolation. Appendix B provides the computer code for obtaining the list of options to test.

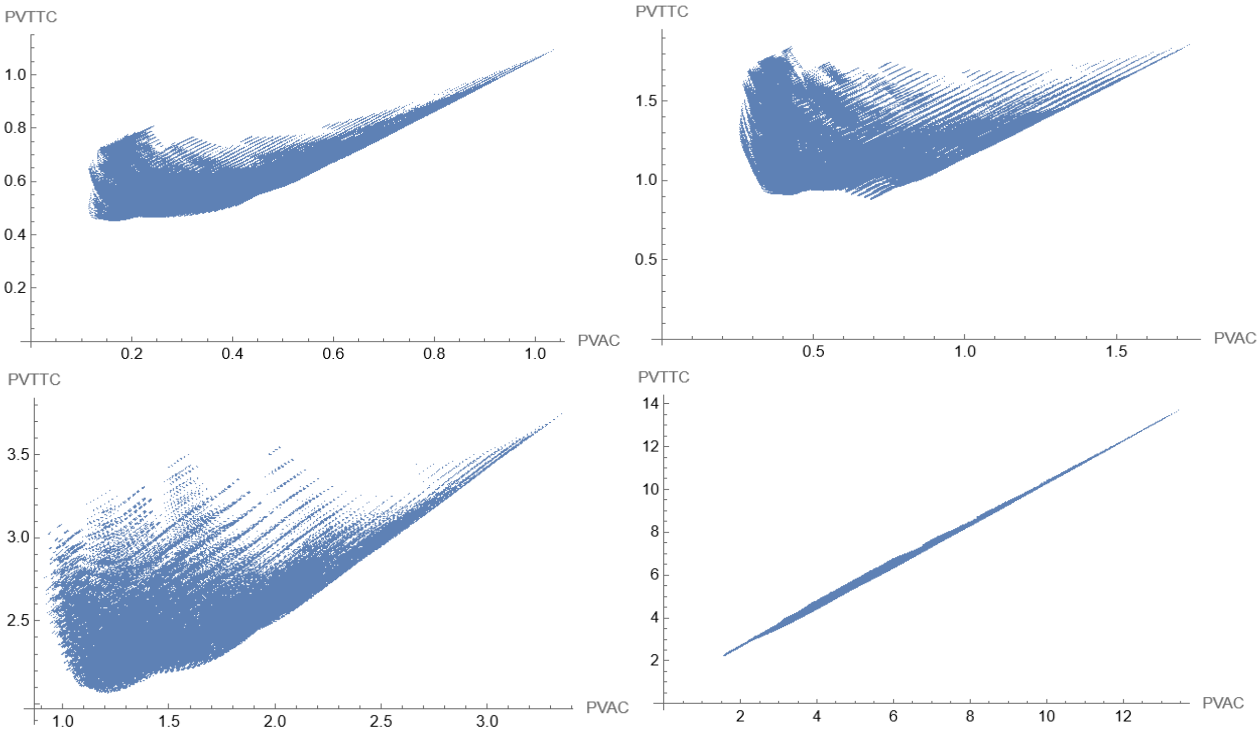
During processing, the model discarded some of the 581,485 solution options for each segment because they violated the technical restrictions. These were mostly options with long intervals of time without treatments during which pavement condition deteriorated severely. The no-treatment option, {{ }, { }}, was not technically feasible for a 40-year analysis period because a pavement neglected for that length time will exceed the maximum allowable roughness level during the period. For segments in very poor condition at the start of the analysis period, for example, a roughness above 6.3 m/km IRI, the technical restrictions made it mandatory to implement a treatment in year one. In such cases, all options for which the first treatment occurred after year one did not need to be tested, saving on model run time. Under the full enumeration approach, having estimated the costs for the full list of options that comply with the technical restrictions, the model simply selected the one with the lowest PVTTC or PVAC value as required.

Figure 5.1 shows plots of PVTTC against PVAC values for all options that did not violate the technical constraints in the model for four segments out the 2034 processed — actual examples of the conceptual illustration in Figure 3.9. Numbers of points shown in each panel range from 300,000 to 570,000. The points with the lowest PVTTC for each PVAC value trace out a U-shaped curve similar to the curve in Figure 3.3, but show that, in practice, there can be irregularities in the U-shape and more than one local minimum. The parallel lines traced out by the points arise where a group of similar options is shifted in time in one-year increments. The segment in the lower-right plot is unusual. It had a low initial roughness level and only required resurfacing treatments during the analysis period. In the lower-right plot, the U-shape is still present in the lower left corner of the mass of points but the left side of the U-shape is limited to a single point north-west of the minimum.

----------

[27] Calculated from $x^a = 10^{a\,Log_{10}(x)}$

**Figure 5.1 Plots of PVTTC against PVAC for all options assessed for four segments**




# 5.5 Case study results

## 5.5.1 Unconstrained total cost minimisation

Only the first 20 years of results are presented here (the 'focus period') because the results for years 20 to 40 are of less interest and were only included in the model to minimise possible distortions of the early-year results from having a finite analysis period.

Table 5.4 summarises the results in percentages of the network treated and spending, in total and for the three main treatment types. Figures 5.2 and 5.3 respectively show the forecast spending and kilometres treated including a breakdown by treatment type, comparing results from the BITRE and ARRB models.

**Table 5.4 Summary of BITRE modelling results: unconstrained optimisation minimising PVTTC**

| Years | Percent of network kilometres treated | | | | Spending ($ millions) | | | |
|---|---|---|---|---|---|---|---|---|
| | **Resurf.** | **RSC** | **Rehab.** | **Total** | **Resurf.** | **RSC** | **Rehab.** | **Total** |
| | | | | **Totals** | | | | |
| **1** | 48 | 4 | 2 | 54 | 103 | 35 | 48 | 186 |
| **1 to 10** | 62 | 8 | 26 | 97 | 132 | 70 | 701 | 903 |
| **11 to 20** | 59 | 17 | 11 | 87 | 85 | 133 | 384 | 602 |
| **1 to 20** | 122 | 26 | 37 | 184 | 217 | 203 | 1,085 | 1,505 |
| | | | | **Annual averages** | | | | |
| **1 to 10** | 6 | 1 | 3 | 10 | 13 | 7 | 70 | 90 |
| **11 to 20** | 6 | 2 | 1 | 9 | 9 | 13 | 38 | 60 |
| **1 to 20** | 6 | 1 | 2 | 9 | 11 | 10 | 54 | 75 |

Notes: Percentages of network kilometres treated in excess of 100% occur where the same road segments are treated more than once over the time period.
Resurf. = resurface, RSC = resurface with shape correction, Rehab. = rehabilitation

The economically optimal solution required the road agency to spend $1505 million over the 20 years, an annual average of $75 million. On average, 9% of the network by length would be treated each year. Spending was higher over years 1–10 than over years 11–20, with average annual spending of $90 million and $60 million respectively. The split of costs between treatment types for the 20 years was 14% for resurfacing, 14% resurfacing with shape correction and 72% rehabilitation. The model rehabilitated just over a quarter of the network during the first 10 years, another tenth during the second 10 years, and very little after year 20 (not shown in Table 5.4), because the network's condition had been brought up to a high standard.

**Figure 5.2 Forecast optimal expenditure without budget constraints**

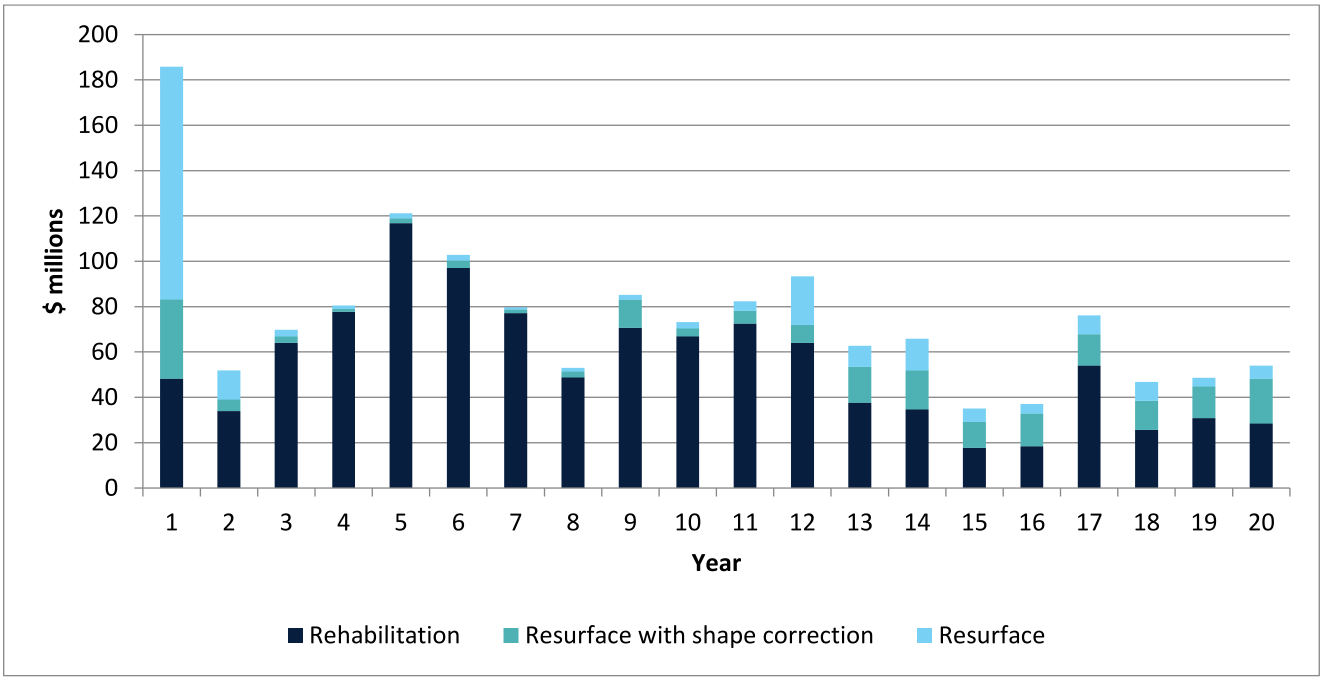
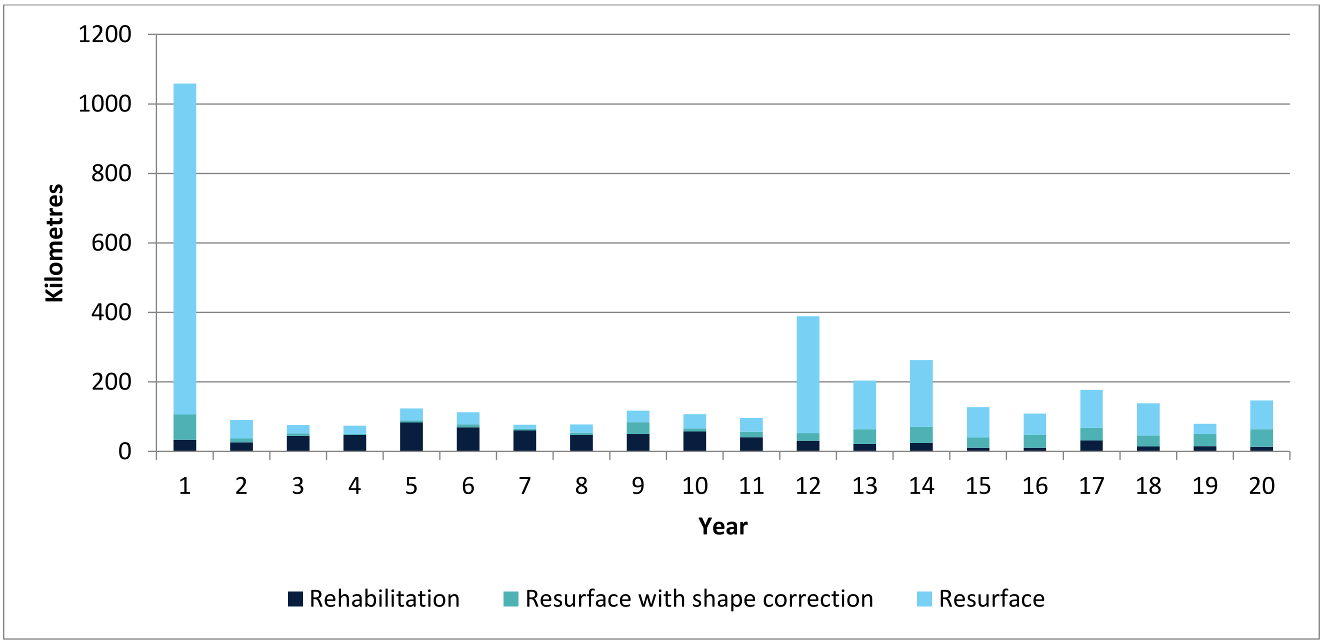



**Figure 5.3 Forecast optimal road lengths treated without budget constraints**

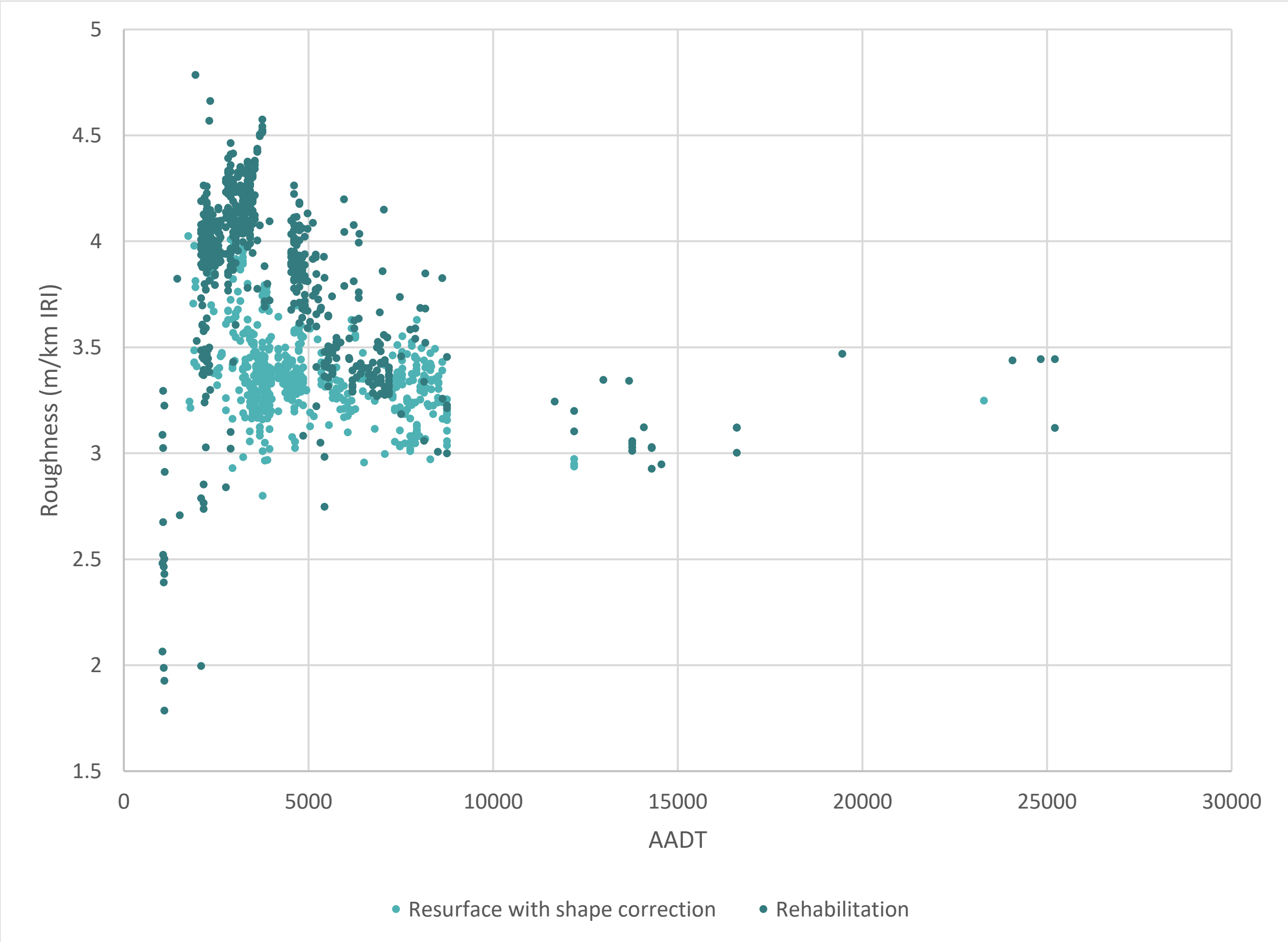



The $186 million amount in the first year to treat 54% of the network suggests there was a substantial 'backlog' of maintenance work that would ideally be undertaken as soon as possible. Of the kilometres treated in year one, 90% were resurfaced (55% by spending) reflecting the fact that, as shown in Table 5.1, 50% of the network by length had cracking of 3% and above. However, it is demonstrated in Chapter 6 that nearly half of the year-one optimal spending could be deferred for a short period at little cost to society.

Figure 5.4 plots roughness just prior to treatment against AADTs as forecast by the BITRE model for all resurface with shape correction and rehabilitation treatments recommended from years 2 to 20.[28] Resurfacing treatments are not shown because they have little impact on roughness. Treatments carried out at the end of year one were excluded because many of them were carried out above roughness levels that would be economically optimal. The expected inverse relationship between AADT and maximum roughness before intervention is evident. The chart shows that for AADTs up to 5000 vehicles per day, the model rarely allowed roughness to rise above 4.5 m/km IRI. For AADTs between 5000 and 9,000 vehicles per day, the maximum was, for the most part, in the range of 3.5 to 4.0 m/km IRI. Above 11,000 vehicles per day, the maximum was 3.5 m/km IRI. Rehabilitation treatments, being more expensive and more effective, tend to be carried out at higher roughness levels, which is to be expected.

Figure 5.4 shows some treatments occurring at very low roughness levels, the lowest being at 1.8 m/km IRI, on low-trafficked roads, which seems difficult to justify on an economic basis. In all such cases, a rehabilitation was triggered by the technical restriction that the pavement strength not be allowed to fall below an adjusted structural number (SNP) of 3.0. Other reasons for treatments at relatively low roughness levels were that rehabilitation costs rise as road condition worsens and resurfacing with shape correction treatments cease to be feasible above specified cracking levels.

**Figure 5.4 Roughness levels at which treatments undertaken plotted against AADT**




Note: AADTs and roughness levels were recorded at the end of the year just before a resurface with shape correction or rehabilitation treatment was undertaken.

----------

[28] The road condition measures in the model could be considered to apply on the 30 December of each year and treatments occur on the 31 December.

## 5.5.2 Initial road condition

As a way of exploring the determinants of maintenance spending needs further, Table 5.4 shows correlation coefficients between fields in the data base at year zero and forecast spending for the first 10 years and 20 years. For each of the 2034 segments, annual spending amounts over the first 10 and first 20 years of the analysis period were summed and divided by segment length in kilometres to remove the effect of segment length on spending.

Correlations were stronger with spending over the first 10 years than over the first 20 years because treatments from years 10 to 20 are further removed in time from the initial pavement condition data. AADT is an exception with a higher correlation coefficient for 20 years, which is understandable as AADT is a persistent property of a road segment not an initial condition that changes with a treatment. Roughness in year zero had the highest correlation coefficients indicating it was the most important influence on spending needs, followed by cracking. The negative correlation with pavement strength is expected because weaker pavements deteriorate faster and are more likely to require treatment. The percentage of heavy vehicles and rut depth have little relationship with spending needs.

**Table 5.5 Correlation coefficients between initial pavement condition variables and forecast spending per kilometre for all 2034 segments**

| | First 10 years | First 20 years |
|---|---|---|
| **AADT** | 0.11 | 0.21 |
| **% Heavy vehicles** | –0.06 | –0.01 |
| **Cracking** | 0.47 | 0.30 |
| **Pavement strength** | –0.30 | –0.27 |
| **Rut depth** | 0.04 | –0.03 |
| **Roughness** | 0.56 | 0.43 |

## 5.5.3 Time intervals between treatments

The time intervals between treatments are of interest because a minimum time interval was set to keep the number of treatment options to a manageable level. Setting aside the time interval between year zero and the first treatment, for each segment, there can be up to four time intervals between subsequent treatments. For example, a segment with five treatments in its optimal solution occurring in the years {2, 10, 20, 29, 38} would have time intervals between treatments of {8, 10, 9, 9} years. A segment with two treatments occurring in the years {15, 35} would have just one time interval between treatments of 20 years. Table 5.6 presents an analysis of time intervals between treatments in the optimal solution based on pavement type and the treatment types at the start and the end of each time interval.

The average time intervals were 11.2 years for surface treatment pavements and 16.1 years for asphalt mix pavements. These are close to the assumed times to crack initiation of 12 years and 16 years respectively. Time intervals leading up to a rehabilitation, the most costly and most effective treatment, which effectively creates a new pavement, were generally longer and could be up to 25 years as deterioration was allowed to progress to a high level just before the rehabilitation. Time intervals following the less effective and less costly resurface with shape correction treatment tended to be shorter. Some treatment sequences occurred rarely or not at all for one pavement type or the other, based on counts of 5, 2 or zero occurrences shown in Table 5.6.

**Table 5.6 Time intervals between treatments**

| | | Surface treatment pavement | | | Asphalt mix pavement | | |
|---|---|---|---|---|---|---|---|
| Start treatment | End treatment | Count | Time interval range (years) | Time interval avg (years) | Count | Time interval range (years) | Time interval avg (years) |
| **Resurf.** | **Resurf.** | 816 | 8 – 16 | 11.7 | 38 | 16 – 20 | 17.6 |
| **Resurf.** | **RSC** | 1150 | 8 – 17 | 12.2 | 110 | 14 – 21 | 18.0 |
| **Resurf.** | **Rehab.** | 138 | 8 – 21 | 13.1 | 34 | 15 – 23 | 19.5 |
| **RSC** | **Resurf.** | 17 | 9 – 13 | 11.5 | 2 | 19 – 19 | 19.0 |
| **RSC** | **RSC** | 1415 | 8 – 17 | 9.7 | 349 | 8 – 21 | 12.3 |
| **RSC** | **Rehab.** | 44 | 8 – 20 | 8.4 | 5 | 8 – 23 | 18.6 |
| **Rehab.** | **Resurf.** | 588 | 10 – 16 | 11.9 | 89 | 16 – 19 | 17.2 |
| **Rehab.** | **RSC** | 0 | na | na | 106 | 17 – 23 | 21.3 |
| **Rehab.** | **Rehab.** | 2 | 18 – 19 | 18.5 | 46 | 19 – 25 | 22.4 |
| **All** | | 4170 | 8 – 21 | 11.2 | 779 | 8 – 25 | 16.1 |

Notes: The table excludes time intervals between year 0 and the first treatment.
Resurf. = resurface, RSC = resurface with shape correction, Rehab. = rehabilitation, avg = average

## 5.5.4 Minimising agency costs subject to standards constraints

The cost–effectiveness analysis approach discussed in Chapters 3 and 4 — minimising the present value of road agency costs subject to minimum road condition constraints — was implemented on the case study data. Maximum roughness constraints were set below the general 6.3 m/km IRI maximum imposed as a technical restriction, and the model was changed to select the option for each segment with the lowest PVAC value.

In order to represent the situation of a severely budget-constrained road agency, the standards were set at the lowest acceptable level for roads on the Australian National Network. The Australian Government's Notes on Administration for Land Transport Infrastructure Projects (DITRD 2013, p. 26) provides a method "to indicate the adequacy of a road's riding quality to meet its transport objectives based on the road's roughness", which calculates boundaries between quality bands expressed in IRI units. The formula to calculate a boundary roughness level is

$$Boundary\ IRI = k\left(\frac{110}{SL}\right)^{0.5}\left[\frac{7.1}{(Car\_AADT + 4 \times HV\_AADT)^{0.11}} + 0.05\right]$$

where

- $SL$ = speed limit in kilometres per hour
- $Car\_AADT$ = AADT for cars
- $HV\_AADT$ = AADT for heavy vehicles
- $k$ is a constant set at
  - 1.0 for the boundary between 'good' and 'mediocre'
  - 1.3 for the boundary between 'mediocre' and 'poor'
  - 1.6 for the boundary between 'poor' and 'very poor'.

The formula accounts for the fact that higher maintenance standards are economically warranted for roads with higher traffic levels, higher proportions of heavy vehicles and higher speed limits. Heavy vehicles are weighted by a factor of four to reflect the higher user costs per vehicle compared with cars. For the purposes of our case study, the boundary between 'poor' and 'very poor' ($k$ = 1.6) was chosen as the upper limit on roughness. All segments in the database were assumed to have speed limits of 100 km/h. The model was set up to reject all treatment options for which roughness in any year after year one exceeds the maximum given by the formula. Year one roughness levels were excluded because they existed just before any year-one treatment could be applied. As traffic levels grow over time, the maximum permitted roughness levels reduce. The 6.3 m/km IRI technical maximum roughness restriction remained in place, but it was redundant. It would apply only for segments with very low AADTs — well below the lowest AADT in the database.

Figure 5.5 was constructed in the same way as for Figure 5.4 except that it also shows, for each data point, the maximum roughness permitted according to the formula given the AADT prevailing at the time of each treatment. The maximum roughness levels plotted show points at different heights for the same AADT level reflecting differences in heavy vehicle proportions.

**Figure 5.5 Cost effectiveness analysis: Roughness levels at which treatments carried out and maximum permitted**

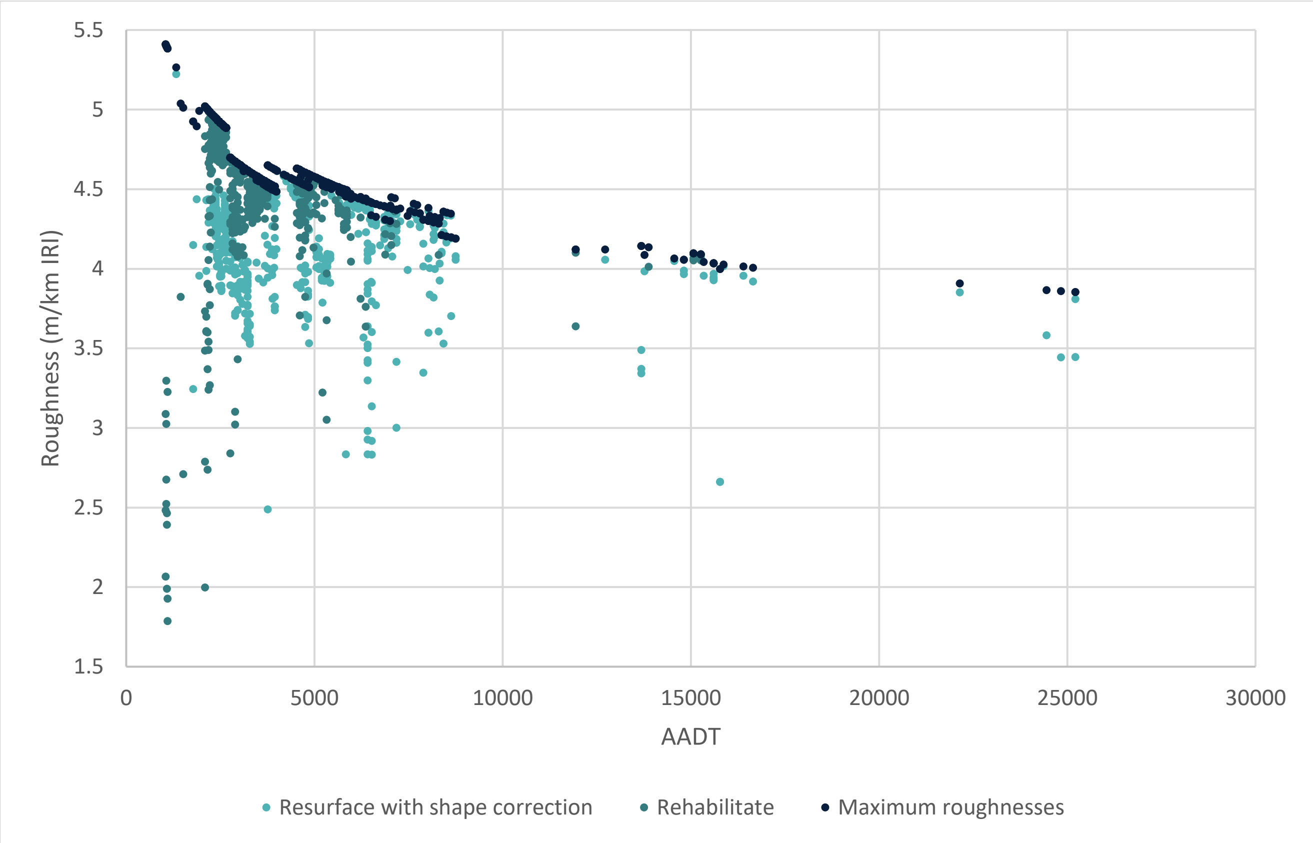



Table 5.7 presents the results in same way as for Table 5.4. Compared with PVTTC minimisation, total undiscounted spending over the 20 years was reduced by $268 million or 18%. The cost saving occurred in the first 10 years, with $283 million saved in the first 10 years followed by a small increase in spending over the second 10 years. Average annual spending reduced from $75 million to $62 million. The total percentage of the network, by length, treated over the 20 years was almost the same.

The main change that saved costs was that the percentage of the network rehabilitated in the first 10 years was halved from 26% to 13%. The percentage split in undiscounted 20-year spending between rehabilitation, shape correction and resurface changed from 72:14:14 to 65:14:21, and kilometres treated changed from 20:14:66 to 14:12:74. By distance, the percentage of the network rehabilitated over the period was 26% compared with 37% for minimising PVTTC.

**Table 5.7 Summary of modelling results: minimising PVAC subject to maximum roughness constraints**

| Years | Percent of network kilometres treated | | | | Spending ($ millions) | | | |
|---|---|---|---|---|---|---|---|---|
| | **Resurf.** | **RSC** | **Rehab.** | **Total** | **Resurf.** | **RSC** | **Rehab.** | **Total** |
| **Totals** | | | | | | | | |
| **1** | 52 | 3 | 1 | 56 | 106 | 30 | 25 | 161 |
| **1 to 10** | 73 | 8 | 13 | 94 | 158 | 69 | 394 | 620 |
| **11 to 20** | 62 | 15 | 13 | 89 | 109 | 101 | 406 | 617 |
| **1 to 20** | 135 | 23 | 26 | 183 | 267 | 170 | 800 | 1237 |
| **Annual averages** | | | | | | | | |
| **1** | 7 | 1 | 1 | 9 | 16 | 7 | 39 | 62 |
| **1 to 10** | 6 | 1 | 1 | 9 | 11 | 10 | 41 | 62 |
| **11 to 20** | 7 | 1 | 1 | 9 | 13 | 9 | 40 | 62 |

Notes: Resurf. = resurface; RSC = resurface with shape correction; Rehab. = rehabilitation
Percentages of network kilometres treated in excess of 100% occur where the same road segments are treated more than once over the time period.

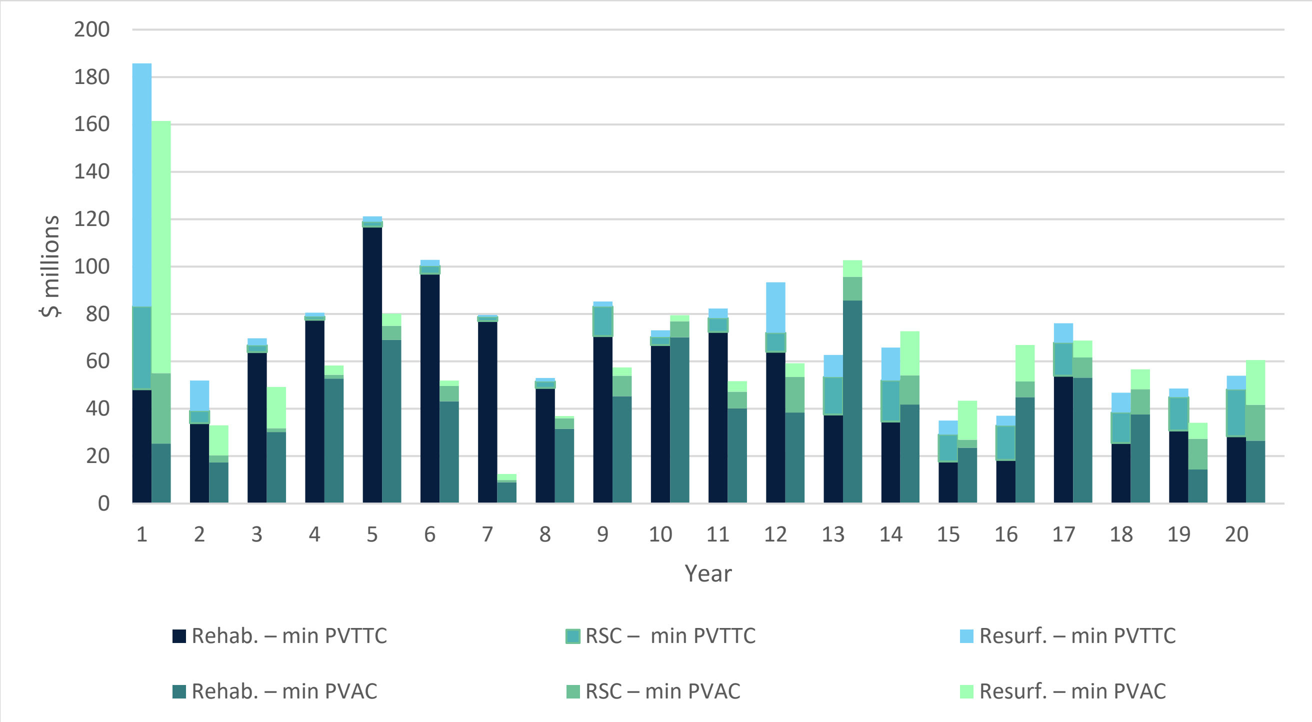
Figures 5.6 and 5.7 respectively compare annual spending and kilometres treated by treatment type for minimising PVTTC (left side of each bar, identical to Figures 5.1 and 5.2) and minimising PVAC subject to minimum standard constraints (right side of each bar). Figure 5.6 shows that changing from PVTTC minimisation to standard-constrained PVAC minimisation reduces spending in all of years one to nine, but leads to higher spending in some years thereafter. The figures confirm that rehabilitation work is deferred under PVAC minimisation.

Figure 5.8 shows distance-weighted average roughness levels for the network under the two modelling approaches. Minimising PVAC subject to a minimum standard constraint gave rise to higher roughness levels, which implies higher road user costs. PVUC was $642 million higher than for the economically optimal unconstrained PVTTC-minimising solution, well above the saving in PVAC of $279 million, a cost to users of $2.30 for each dollar of agency cost saved. The cost to society of the sub-optimal solution was $363 million (= $642 million higher PVUC minus $279 million lower PVAC). Had the same cost saving to the road agency been achieved by minimising PVTTC subject to a present value budget constraint, the cost to society would have been $184 million (= $463 million higher PVUC minus $279 million lower PVAC) because the exogenously-set maximum roughness levels under the cost-effectiveness analysis approach were are not optimal from the point of the view of society.[29] The cost-effectiveness approach is less economically efficient than unconstrained PVTTC minimisation both on account of the below-optimal overall level of spending and the minimum acceptable road condition constraints.

----------

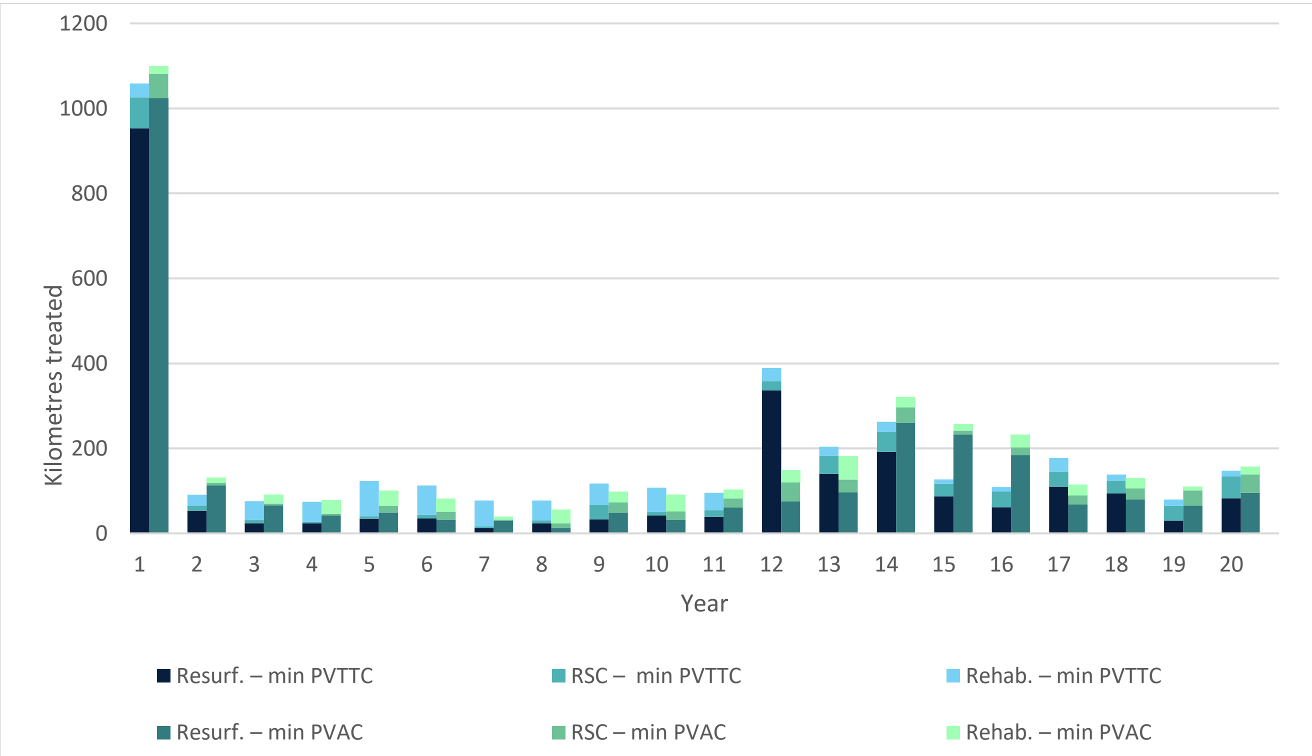
[29] This was estimated by minimising weighted PVTTC, the methodology illustrated in the next section, and interpolating between target MBCRs of 3.0 and 3.5.

**Figure 5.6 Forecast optimal expenditure minimising PVTTC and standard-constrained PVAC**




Notes: The left bar for each year is for minimising PVTTC, identical to the data shown in Figure 5.1. The right bar is for minimising PVAC subject to the maximum roughness constraint.
Rehab. = rehabilitation, RSC = resurface with shape correction, Resurf. = resurface

**Figure 5.7 Forecast optimal kilometres treated minimising PVTTC and standard-constrained PVAC**




Notes: The left bar for each year is for minimising PVTTC, identical to the data shown in Figure 5.1. The right bar is for minimising PVAC subject to the maximum roughness constraint.
Rehab. = rehabilitation, RSC = resurface with shape correction, Resurf. = resurface

**Figure 5.8 Length-weighted average roughness minimising PVTTC and PVAC subject to conditions constraints**


Note: Starting the vertical axis at 2.0 instead of zero accentuates the difference between the two curves.

## 5.5.5 **MBCRs above one**

Chapter 3 discussed minimising PVTTC subject to a budget constraint expressed as a present value, that is, a maximum allowable value for PVAC. For a network of *n* segments with index *i*, the optimisation problem becomes

> Minimise $\sum_{i=1}^{n} PVTTC_i$ subject to $\sum_{i=1}^{n} PVAC_i \leq B$ where *B* is the present value budget constraint for the network.

A simple way to implement this is to set up the model to find the option that minimises of *PVUC + weight × PVAC* for each individual segment and adjust the weight, re-running the model as necessary, to produce the desired PVAC value. It was shown in Chapter 3 that the value of the weight is the MBCR, that is, the benefit from increasing PVAC by one dollar.

For each segment, the model found all technically feasible solutions from the possible 581 485 treatment timing and type combinations. From the list of feasible solutions, the model selected the option with the lowest PVTTC, which is the best overall option in the absence of any budget constraints. While the list for the segment being processed was held in the computer's memory, it was a simple matter to also extract from the list, the options with the lowest values for *PVUC + weight × PVAC* for a selection of specified weights. This was done for weights or MBCRs ranging from 1.5 to 25.

Model results were also extracted for minimising PVAC. In theory, minimising PVAC with no constraints should lead to zero maintenance activity. However, the technical constraints in the model and the increasing cost of treatments and depreciation as road condition falls, gave rise to a basic maintenance scenario. The MBCR is infinite due to PVUC having a zero weighting.

Model outputs are summarised in Table 5.8 with PVUC normalised to zero for the optimum with MBCR equal to one. Regardless of the MBCR, first year spending was always well above annual averages. Comparison of average spending over the two 10-year periods shows that raising the MBCR delays spending.

Figure 5.9 shows annual maintenance expenditures for the first 20 years for MBCRs of 1, 10 and 20. Imposition of a present value budget constraint has a limited impact on warranted first-year expenditure but does significantly delay spending. Annual spending was higher for MBCR = 1 up to year 15, after which spending was higher with MBCR = 10 or 20 for all but year 17.

**Table 5.8 Model results with increasing MBCRs compared**

*($ millions)*

| MBCR | PVTTC | PVAC | PVUC | AC 1 | Average AC 1–10 | Average AC 10–20 | Average AC 1–20 | Correlation AC 1–20 |
|---|---|---|---|---|---|---|---|---|
| **1** | 1902 | 1902 | 0 | 186 | 90 | 60 | 75 | 1.00 |
| **1.5** | 1933 | 1747 | 186 | 181 | 77 | 62 | 69 | 0.76 |
| **2** | 1970 | 1696 | 274 | 175 | 71 | 63 | 67 | 0.69 |
| **2.5** | 2015 | 1660 | 355 | 169 | 67 | 66 | 66 | 0.53 |
| **5** | 2176 | 1594 | 582 | 156 | 57 | 52 | 55 | 0.59 |
| **7.5** | 2282 | 1573 | 709 | 161 | 50 | 54 | 52 | 0.48 |
| **10** | 2340 | 1565 | 775 | 163 | 48 | 56 | 52 | 0.43 |
| **12.5** | 2396 | 1560 | 836 | 167 | 45 | 59 | 52 | 0.38 |
| **15** | 2443 | 1556 | 887 | 166 | 45 | 59 | 52 | 0.36 |
| **17.5** | 2488 | 1553 | 935 | 166 | 44 | 59 | 51 | 0.35 |
| **20** | 2520 | 1551 | 968 | 165 | 44 | 59 | 51 | 0.35 |
| **25** | 2562 | 1549 | 1013 | 166 | 43 | 59 | 51 | 0.35 |
| **Infinite** | 2718 | 1546 | 1171 | 165 | 42 | 55 | 49 | 0.38 |

Notes: PVUC has been normalised to zero with PVTTC = PVAC + PVUC.
AC 1 = road agency cost in year one.
AC *x* – *y* = annual average of undiscounted road agency costs for the years *x* to *y*.
AC 1–20 correlation is the correlation coefficient between annual road agency costs over years 1 to 20 for MBCR = 1 (unconstrained) and the MBCR for the row concerned. It is a measure of the degree to which the spending profile differs from the optimal scenario.

Figures 5.10 and 5.11 plot the values of PVTTC, PVAC and PVUC, with PVUC normalised to zero at the unconstrained optimum with an MBCR of one. Figure 5.11 is equivalent to the portion of Figure 3.3 in Chapter 3 to the left of the optimum. The results show the cost to society of restricting maintenance spending as the additional costs to road users outweighs the saving in agency costs, such that PVTTC was $660 million above the unconstrained optimum when the MBCR reaches 25.

When applied to an individual segment, the discrete nature of the treatment types and implementation years means that the same option will have the lowest weighted PVTTC over a range of values for the weights. However, for a network of segments taken together, the relationships between the present values of costs summed over all segments approximate a smooth curve.

**Figure 5.9 Annual warranted spending with MBCR values of 1, 10 and 20**

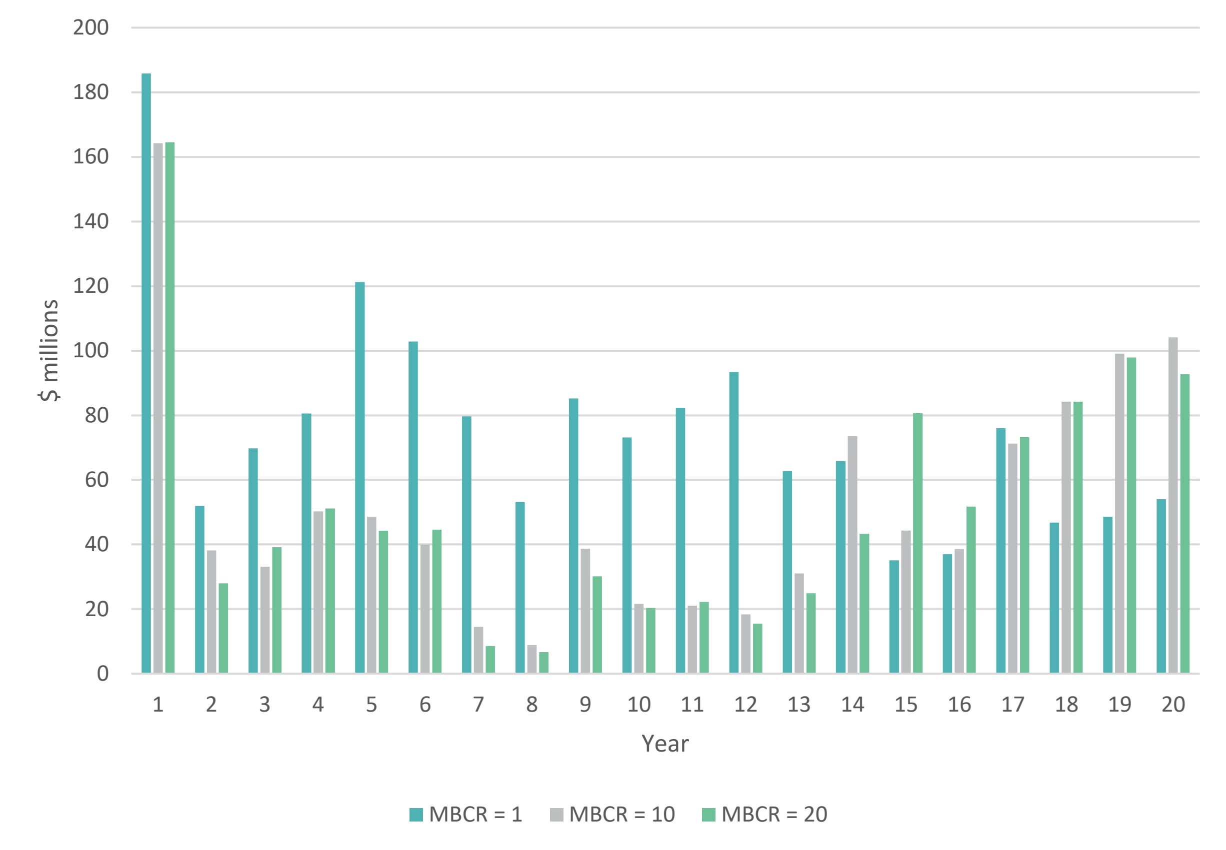



**Figure 5.10 Present values of costs with different MBCR values plotted against MBCR**

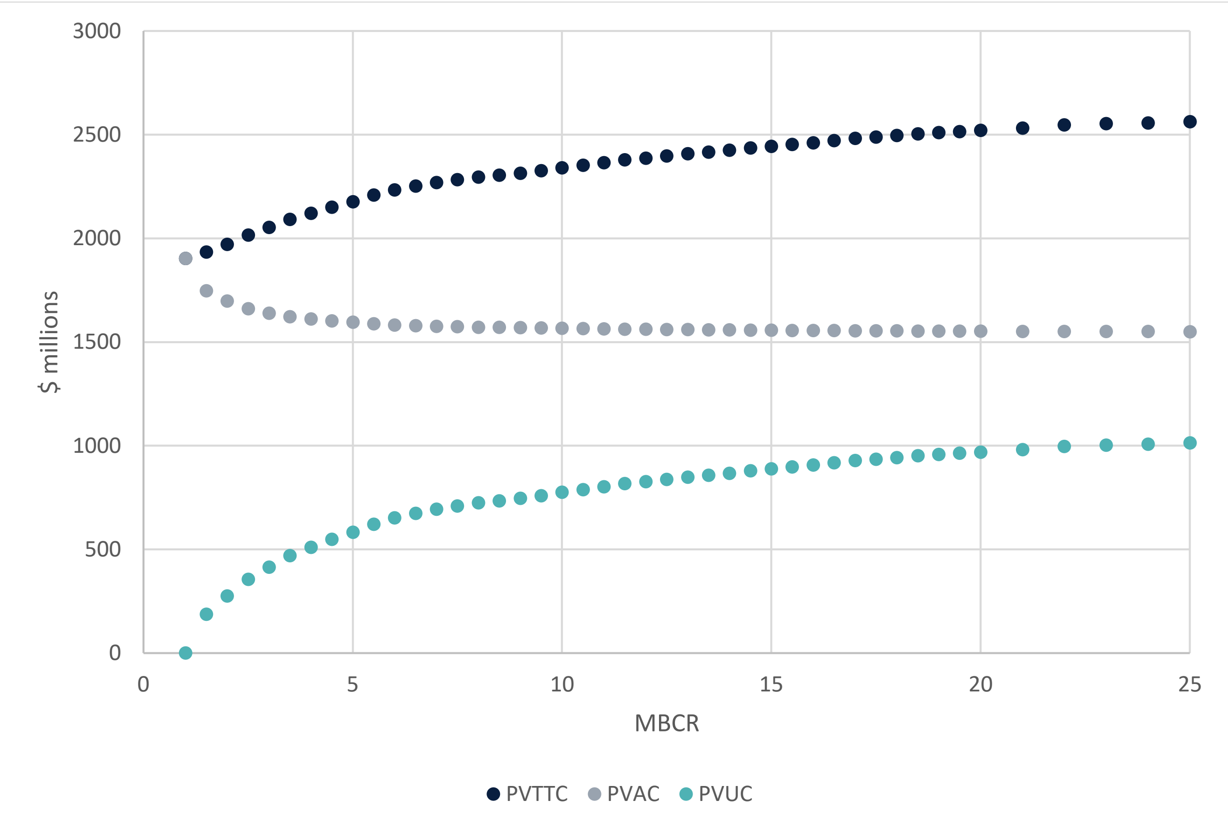
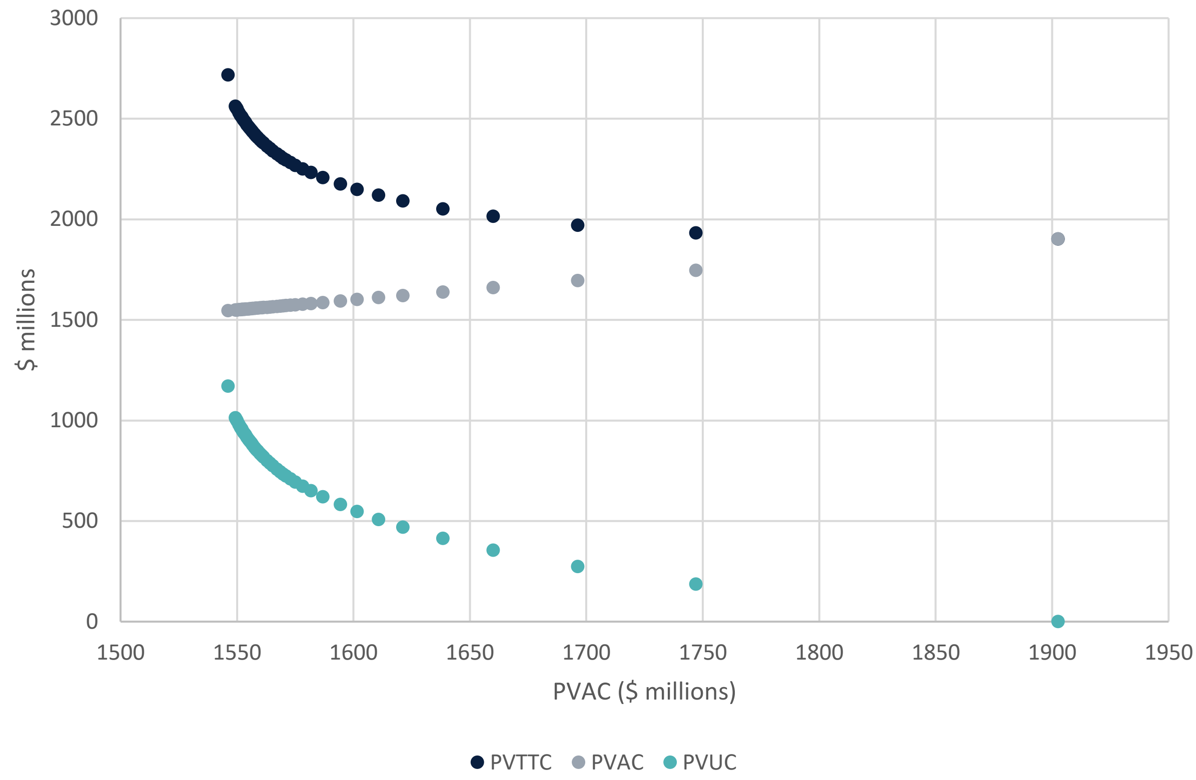


**Figure 5.11 Present values of costs with different MBCR values plotted against PVAC**




Note: Figure 5.11 contains an additional data observation not shown in Figures 5.10 and 5.12. The leftmost point, for a PVAC value of $1 546 million, is for an infinite MBCR found by minimising PVAC subject only to the technical constraints in the model.

**Figure 5.12 MBCR plotted against PVAC saved**

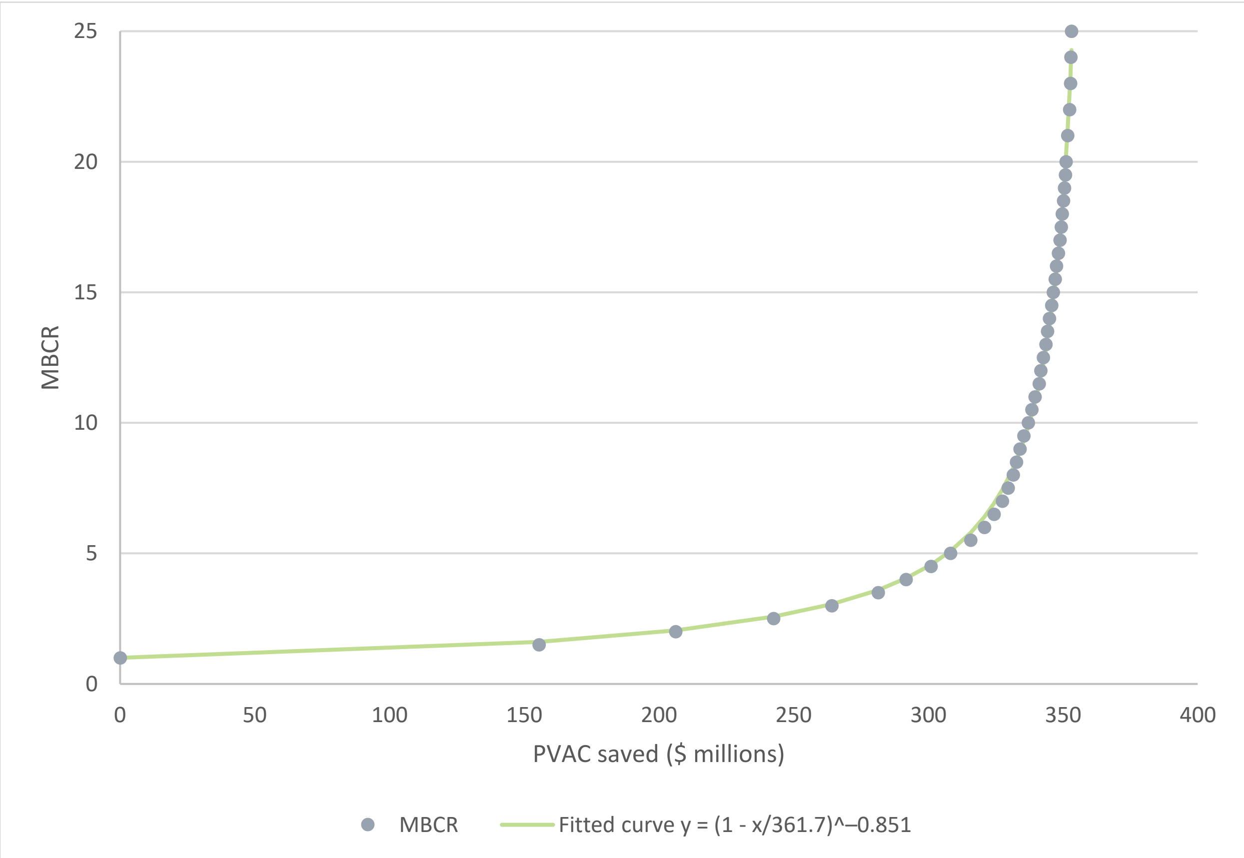


Figure 5.12 displays the outputs in a telling way. The horizontal axis is the amount by which PVAC was reduced compared with the unconstrained optimal value where the MBCR is one, in other words, the cost saving to the road agency. MBCR is plotted on the vertical axis. The first step, raising the MBCR from 1.0 to 1.5 reduced PVAC by $155 million, an 8% saving. The next step, raising the MBCR from 1.5 to 2.0 saved $51 million or 2.7%. Thereafter, the rise in the MBCR for each additional dollar reduction in PVAC became extremely steep. The rise was more than exponential. The fitted curve asymptotically approaches a saving of $362 million, but the model was unable to go beyond a saving of $356 million, or 19%, without relaxing the technical constraints within the model. The finding that MBCRs rise slowly at first as spending is reduced and then increase rapidly as spending is further restricted is repeated for annual budget constraints in the next chapter.

The discussion of incremental BCRs in Chapter 3 provided formulas for incremental BCRs between two points on the minimum PVTTC curve, one of which is

$$IBCR = -\frac{PVTTC_2 - PVTTC_1}{PVAC_2 - PVAC_1} + 1$$

Table 5.9 shows changes in present values and IBCRs calculated between the points where spending for various MBCRs has been estimated. As expected, the IBCRs are always between the MBCRs for the upper and lower points.

**Table 5.9 Incremental BCRs in between marginal BCRs**

| ∆PVTTC ($ millions) | ∆PVAC ($ millions) | ∆PVUC ($ millions) | Lower MBCR | IBCR | Upper MBCR |
|---|---|---|---|---|---|
| –617 | 351 | –968 | 1 | 2.8 | 20 |
| | | | | | |
| –438 | 337 | –775 | 1 | 2.3 | 10 |
| –180 | 14 | –194 | 10 | 13.8 | 20 |
| | | | | | |
| –112 | 243 | –355 | 1 | 1.5 | 2.5 |
| –162 | 66 | –227 | 2.5 | 3.5 | 5 |
| –106 | 21 | –127 | 5 | 5.9 | 7.5 |
| –58 | 7 | –65 | 7.5 | 8.7 | 10 |
| –56 | 6 | –61 | 10 | 11.1 | 12.5 |
| –48 | 4 | –51 | 12.5 | 13.7 | 15 |
| –44 | 3 | –47 | 15 | 16.2 | 17.5 |
| –32 | 2 | –34 | 17.5 | 18.6 | 20 |
| -43 | 2 | -45 | 20 | 21.9 | 25 |

## 5.6 Sensitivity tests

Table 5.10 summarises results from a variety of sensitivity tests in which PVTTC was minimised after changes to the model or to the data. With the exception of the last column, all results are presented as differences from the central scenario result. The columns are

1. PVTTC over the 40-year analysis period including depreciation and safety costs
2. PVAC including depreciation
3. PVUC including safety costs
4. UC 1–20 — undiscounted sum of annual user costs (including safety) over years 1 to 20 — a measure of the impact on users

5. AC 1 — road agency cost (spending) in year one
6. AC 1–10 — undiscounted sum of annual road agency costs (spending) over years 1 to 10
7. AC 10–20 — undiscounted sum of annual road agency costs (spending) over years 10 to 20.
    - Columns 6 and 7 indicate how spending shifts from the first half to the second half of the 20-year focus period.
8. AC 1–20 — undiscounted sum of annual road agency costs (spending) over years 10 to 20.
    - Columns 6 and 7 sum to column 8 with small differences due to rounding.
9. AC 1–20 correlation — correlation coefficient between annual road agency costs compared with the central scenario over years 1 to 20. It is a measure of the degree to which the sensitivity test altered the spending profile compared with the central scenario.

*Discount rate 7%:* Raising the discount rate reduced agency spending overall (6% reduction over the 20 years) at the expense of users and pushed spending into the future. From the correlation coefficients column, it produced the largest rearrangement of the annual spending profile.

*Pavement strength ±20%:* The design adjusted structural numbers were multiplied by 1.2 and 0.8. As the design pavement strength for each segment was assumed to apply when the pavement age was zero, each altered design pavement strength was projected forward to the first year of the analysis period using equation 2.2. The change in pavement strengths had little effect on user costs, but a significant impact on agency costs (–14% for stronger payments and +23% for weaker pavements over the 20 years). The weaker pavement sensitivity test caused large increases in agency costs during the early years. Once the pavements were rehabilitated to bring them up to design standard, the sensitivity test on initial design strengths had no effect, hence the smaller impact during the later years.

*User costs without safety:* The effect of omitting safety from user costs was tested because most maintenance optimisation studies do not take account of safety costs. Optimal treatments selected were those with minimum PVTTC without safety costs, but the present values reported in Table 5.10 include safety costs in order to show the cost of the resulting sub-optimal solution. PVTTC was $26 million higher because omitting safety leads to a worse outcome from the point of view of society. Over the 20-year focus period, undiscounted user costs were $123 million higher compared with a saving to the road agency of $79 million due to pushing maintenance spending into the future.

*No depreciation:* A few studies in the literature have no residual value or depreciation at the end of the analysis period, relying instead on a long analysis period to allow for the far future. The optimal treatments selected by the model were those with minimum PVTTC without depreciation, but the present values reported in Table 5.10 include depreciation in order to show the cost of the resulting sub-optimal solution. The sensitivity test shows that, with a 40-year analysis period and 4% discount rate, the effects of omitting depreciation were fairly small, but there was some disadvantage to road users over the 20-year focus period. The impact would be much greater with shorter analysis periods, though this would be offset if the discount rate was higher.

**Table 5.10 Results of sensitivity tests**

*($ millions deviation from the central scenario except for column 9)*

| Sensitivity test | PVTTC | PVAC | PVUC | UC 1–20 | AC 1 | AC 1–10 | AC 10–20 | AC 1–20 | AC 1–20 correlation |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| **Central scenario** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.00 |
| **Discount rate 7%** | –1612 | –847 | –765 | 207 | –8 | –225 | 140 | –85 | 0.46 |
| **Pavement strength +20%** | –209 | –190 | –19 | 37 | –1 | –199 | 11 | –188 | 0.77 |
| **Pavement strength –20%** | 361 | 377 | –16 | 37 | 61 | 237 | 209 | 447 | 0.74 |
| **User costs without safety**[a] | 26 | –79 | 105 | 123 | –12 | –75 | 42 | –33 | 0.81 |
| **No depreciation**[a] | 50 | –8 | 58 | –8 | 12 | 11 | –11 | 0 | 0.95 |
| **Analysis period 30 years**[b] | na | na | na | –3 | 13 | 28 | –144 | –116 | 0.90 |
| **Analysis period 20 years**[b] | na | na | na | 168 | 28 | –150 | –341 | –491 | 0.84 |
| **Only even years > 22** | 2 | 4 | –3 | –2 | 10 | 1 | 4 | 5 | 0.98 |
| **Every third year after 22** | 6 | 0 | 6 | 1 | 0 | –4 | 22 | 18 | 0.94 |
| **Only years divisible by 5 >22** | 15 | –1 | 16 | –1 | 1 | –13 | 12 | –1 | 0.88 |
| **Minimum interval between treatments 11 years** | 21 | –10 | 31 | 0 | 0 | –1 | –26 | –26 | 0.79 |

Notes: AC 1 = road agency cost in year one.

AC *x* – *y* = annual average of undiscounted road agency costs for the years *x* to *y*.

AC 1–20 correlation is the correlation coefficient between annual road agency costs over years 1 to 20 for MBCR = 1 (unconstrained) and the MBCR for the row concerned. It is a measure of the degree to which the spending profile differs from the optimal scenario.

a. For the sensitivity tests excluding safety and depreciation, the optimal selection of treatment options changed, but the present values reported in the table include safety and depreciation in order to show the cost to society of the resulting sub-optimal solution.

b. Changes in present values for shortening the analysis period are not reported because present values over different analysis periods are not comparable.

The last six sensitivity tests explored the effects of ways to reduce the number of options and hence model run times. Table 5.11 details these sensitivity tests including the impact on the number of options to be tested.

*Shorter analysis period:* As evident in Table 5.11, shortening the analysis period is highly effective in reducing the number of treatment options. With a 30-year analysis period and an eight-year minimum interval between treatments, a maximum of four treatments (for example treatments in years 1, 9, 17 and 25) can be undertaken within the analysis period. Further reducing the analysis period to 20 years, a maximum of three treatments (for example treatments in years 1, 9 and 17) can be undertaken. Table 5.10 shows that the impact on results was significant, particularly for the 20-year analysis period. For the 30-year analysis period, there was little effect on undiscounted user costs, while undiscounted agency costs for the first 20 years were 8% lower. For the 20-year analysis period, undiscounted user costs were significantly higher and undiscounted agency costs down by 48%. The reduction in agency costs means the network was in poorer condition at the end of year 20 compared with the 40-year analysis period. Shortening the analysis period places greater reliance on the depreciation formula to approximate the PVTTC for years after the end of the analysis period. Improving residual value or depreciation formulas might be a topic for future research but the degree of improvement possible may be limited. A higher discount rate would also make shorter analysis periods less distorting.

*Skipping analysis years:* For the three sensitivity tests in which some of the later years of the analysis period were skipped, years 1 to 22 were retained. Skipping year 21 would push some year 21 maintenance spending into year 20, which is within the focus period for the study. For the 'only even years >22' sensitivity test, treatment timing combinations such as {5, 19, 30} would be assessed but not {5, 19, 31}. For the 'only years divisible by 5 >22' test, treatment timing combinations such as {2, 14, 25, 40} would be assessed but not {2, 14, 26, 40}. The results in Table 5.11 indicate that omitting some years in the latter part of the analysis period has only a limited effect on the model's results, though the declining correlation coefficients as more years are skipped suggests that dropping a larger number of years caused some changes to the spending profile during the first 20 years.

**Table 5.11 Details of sensitivity tests of ways to reduce model run times**

| Sensitivity test | Possible treatment years | Minimum years between treatments | Number of options | Number of options / 581 485 % |
|---|---|---|---|---|
| **Case study model** | 1–40 | 8 | 581 485 | 100.0% |
| **Analysis period 30 years** | 1–30 | 8 | 27 694 | 4.8% |
| **Analysis period 20 years** | 1–20 | 8 | 1 303 | 0.2% |
| **Only even years > 22** | 1–22, 24, 26, 28, 30, 32, 34, 36, 38, 40 | 8 | 208 831 | 35.9% |
| **Every third year after 22** | 1–22, 25, 28, 31, 34, 37 ,40 | 8 | 97 411 | 16.8% |
| **Only years divisible by 5, >22** | 1–22, 25, 30, 35, 40 | 8 | 58 120 | 10.0% |
| **Minimum interval between treatments 11 years** | 1–40 | 11 | 51 826 | 8.9% |

*Minimum interval between treatments 11 years:* In this sensitivity test, the minimum time interval permitted between treatments was increased from eight to 11 years, just below the 12-year crack initiation period (the period of time before cracking starts rise) for surface treatment pavements. An example of a treatment timing combination with 11-year intervals is {1, 12, 23, 34}. The maximum number of treatments that could be fitted within the analysis period reduces from 5 to 4. The results in Table 5.11 show that the impact on user costs and spending was relatively small, but the correlation coefficient indicates there was some significant rearrangement of the timings of spending in the first 20 years.

Conclusions from the sensitivity tests are

- Higher discount rates reduce and delay maintenance spending but at a higher cost to road users.
- Pavement strengths at the start of the analysis period can have a large impact on road agency costs in total and in timing.
- Excluding safety costs from the analysis disadvantages road users with a small reduction in road agency costs.
- Shorter analysis periods relative to the focus period can distort results because depreciation estimates are highly approximate predictors of PVTTC beyond the end of the analysis period. It is better to reduce reliance on depreciation estimates by having a longer analysis period.
- If the time to run the model needs to be reduced, it is better to skip some years in the latter part of the analysis period and to increase the minimum time interval between treatments (but not more than the number of years before crack initiation) than to shorten the analysis period.

## 5.7 Conclusion

For the report's case study without annual budget constraints, full enumeration of all possible solutions was chosen as the optimisation method to guarantee finding the optimal solution. The absence of annual budget constraints meant that each segment could be optimised in isolation from the other segments. But even then, with a 40-year analysis period and three treatment types, the number of possible solutions was unmanageable. As cracking does not start to occur until the pavement's surface age reaches 12 years or more, it was considered unlikely that the economically optimal solution would involve any periodic maintenance treatments with a time interval between them of much below 12 years. Eight years was chosen as the minimum time gap between treatments reducing the number of possible solutions for each segment below 600,000.

In the absence of annual budget constraints, minimising PVTTC or PVAC subject to maximum roughness constraints usually leads to a large spike in spending in the first year and spending in subsequent years can be highly uneven. Hence, there is therefore a need to optimise subject to annual budget constraints, which is addressed in the next chapter.

As discussed in Chapter 3, imposing budget constraints in the form of present values of road agency costs, while unrealistic, is straightforward to implement provided one starts by setting an MBCR. Setting MBCRs in small steps from 1.5 to 25 showed that a road agency can reduce spending at relatively little cost to users up to a point. The MBCR rises gently at first as spending is reduced. However, as the budget is further reduced, the rise in MBCR, and hence the cost of society of each dollar of PVAC saved, rapidly accelerates until reaching the minimum possible PVAC value within the technical constraints of the model. This finding is repeated in the next chapter for annual budget constraints. Constraints imposing small-to-moderate spending cuts can be imposed at little cost to society, but the cost rapidly rises for tighter constraints.

If computer run times for the model need to be reduced, the sensitivity tests showed that it is better to retain the longer 40-year analysis period and skip some years in the later part of the period (that is, not to test options with treatments in the skipped years) than to shorten the analysis period to 30 or 20 years.

# 6. Case study with annual budget constraints

## Summary

Optimising subject to annual budget constraints requires all segments to be considered together. A four-stage optimisation approach was developed for the case study. The stages were

1. Full enumeration of options for all segments, already discussed in Chapter 5
2. Elimination of 'dominated' options that could not possibly appear in an optimal solution subject to annual budget constraints because there is a better option
3. Selection of the option for each segment that minimises the objective function (PVTTC or PVAC) plus the road agency spending in each budget-constrained year times a 'penalty factor'. A higher penalty factor for a given year discourages selection of options with treatments in that year. The penalty factors are adjusted so that the annual budget constraints are met
4. Fine adjustment of the solution by allowing a genetic algorithm to select from the available options to minimise the objective function subject to the budget constraints. This mitigates the limitation of the penalty method that it treats a discrete problem as if it were continuous. Treatments are shifted between years to take advantage of any gaps between forecast spending and annual budget constraints.

Provided the penalty factors are at the lowest possible values to achieve the solution, MBCRs can be obtained for each year using the formula from Chapter 3, $MBCR_t = (1 + r)^t \lambda_t + 1$, where $\lambda_t$ is the penalty factor for year *t*.

Case study results with annual budget constraints for the first 10 years and the first 20 years showed that

- The required penalty factors and hence MBCRs were highest for the first year when the demand for funds is greatest and declined thereafter.
- A substantial proportion of the year-one maintenance backlog could be deferred at little cost.
- Penalty factors increased at an increasing rate as annual budget constraints were tightened.
- As constraints were tightened, PVAC fell initially, and then rose as the additional costs in later years to recover from the underspending in early years predominated.
- There were limits on how much spending could be constrained due to the technical constraints imposed in the model.

A simple triaging method for maintenance treatments using only a penalty factor for year one, was demonstrated to work satisfactorily for modest budget constraints, but not for tight constraints.

The first-year spending backlog is not a good measure of the maintenance deficit because a large part of it is not urgent. Maintenance deficits are better measured by comparing either the 'sustainable' level of annual spending or average annual forecast spending with current or forecast spending. A 'sustainable' level of spending could be defined as one where there is no jump in optimal spending just following the constrained period, or the jump is not so large that it cannot be caught up by continued spending at the sustainable level in subsequent years. Annual MBCRs can also serve as measure the maintenance deficit and are directly comparable with the BCRs for capital spending.

The 'equivalent interest rate for deferred maintenance' is proposed as a way to convey to decision makers the costs of deferring maintenance spending. Saving funds in the short term in exchange for spending more in later years to repair the damage done is like borrowing money that has to be repaid later. However, it can be a very expensive way to borrow.

## 6.1 Introduction

The case study in Chapter 5 showed that economically optimal maintenance needs forecast by a model in the absence of budget constraints can be very uneven from year to year, especially in the first year of the analysis period, when there is usually a large spike or backlog in spending needs. Present value budget constraints do not rectify the problem. Imposition of annual budget constraints is therefore needed to keep treatment recommendations from an optimisation model within financial and physical resource constraints.

Annual budget constraints give rise to a new set of analytical difficulties because maintenance needs for each segment can no longer be optimised in isolation from the other segments. With a binding annual budget constraint, spending on one segment in any year comes at the expense of spending on other segments in that year. It is therefore necessary to model all segments together. But this greatly magnifies the curse of dimensionality. As discussed in Chapter 5, with a 40-year analysis period, four treatment types including the null treatment, and a minimum of an eight-year time interval permitted between treatments, there are up to 581,485 options for each segment. The total number of possible solutions for the 2034 segments in our case study database is therefore up to $581{,}485^{2034} \approx 10^{11{,}725}$. This is less than the $4^{40\times 2023} \approx 10^{48{,}984}$ solutions (from the formula in Section 4.5) where there is no minimum time interval between treatments.

This chapter continues the case study of Chapter 5 by finding optional solutions with annual budget constraints lasting for the first 10 and 20 years of the analysis period. Results are presented both for PVTTC minimisation and PVAC minimisation subject to road condition constraints. The dimensionality problem is addressed though a four-stage optimisation approach. MBCR values are derived for individual years showing the benefit to society from a small increase in the budget for a given year, holding budgets in other years constant.

A simple method for triaging year-one treatments, that is, identifying year-one treatments in the unconstrained optimal solution that can be deferred at low cost is presented, but it works only for moderate reductions in year-one spending.

The chapter ends with a discussion of ways to measure the size of 'maintenance deficits'.

# 6.2 Optimisation method for annual budget constraints

## 6.2.1 **Overview**

Typically, in the maintenance optimisation literature, a uniform annual budget constraint is set for a network of segments for the first several years, then no constraints thereafter. Periods of 10 and 20 years were chosen for annual budget constraints rather than shorter periods such as the five years in Archondo-Callao (2008) because of the large bulge in delayed spending in the years immediately following the constrained years. With 10-year constraints, there can be a bulge in year 11, but it is further into the future compared with year six in the case of five years of constraints.

For annual budget constraints, the network optimisation problem is to find the maintenance spending in each year *t* on each segment, $c_{it}$, to minimise network PVTTC

$$\sum_{i=1}^{n} PVTTC_i(c_{i1}, c_{i2}, \ldots, c_{im}, c_{im+1}, \ldots)$$

subject to annual budget constraints for the first *m* years, $\sum_{i=1}^{n} c_{it} \leq B_t$ for all *t* from 1 to *m*.

Initial attempts to minimise PVTTC subject to annual budget constraints for 10 years using a genetic algorithm were unsuccessful, despite using a smaller database for the case study network of 573 'strategic segments' developed by ARRB. Faced with an astronomically large number of choices and a starting point well away from the optimum, the genetic algorithm experienced diminishing returns in terms of reductions in PVTTC for each step without having made much progress towards meeting the constraints.

The optimisation method employed for annual budget constraints drew on two techniques discussed in the literature review in Chapter 4.

- Torres-Machí et al. (2014) and Medury and Madanat (2014) discussed two-stage approaches. Two-stage approaches deal with individual segments in isolation in the first stage, developing a list of solutions for each segment in order of priority. In the second stage, they move down the priority list for each segment as required to meet the budget constraints.

- For imposing budget constraints on a list of options, the Lagrange multiplier approach in Sathaye and Madanat (2011) was employed. It is equivalent to the 'penalty method' used with genetic algorithms.

The optimisation approach in this chapter consists of four stages.

1. Full enumeration of all treatment options for all segments subject to a minimum time interval between periodic maintenance treatments (covered in Chapter 5)
2. For each segment, eliminate all 'dominated' options to reduce the numbers of options to manageable levels
3. Find the best possible solution using the penalty method
4. Allow a genetic algorithm to improve on the solution by giving it complete freedom to change options.

### 6.2.2 Stage 2: Eliminating dominated options

In the first stage of the optimisation process using full enumeration, covered in Chapter 5, road condition, maintenance expenditures and user costs were projected forward for 40 years for up to 581,485 treatment options for all 2024 segments. Exclusion of options that violate the technical constraints in the model (for example, maximum permitted roughness and minimum permitted pavement strength) reduced the total number of options for each segment by varying amounts. The largest reductions were for segments that had to be treated in year one because of technical constraints in the model. In such cases, options where the first treatment occurred in year two and onwards did not need to be considered, leaving 154,677 options to be assessed. But only 11 segments out of the 2034 fell into this category.

Stage 2 dramatically reduced the size of the list of options by eliminating all options that were 'dominated' in the sense that they could not possibly appear in the optimal solution subject to budget constraints because a better option exists. This was be done for each segment in isolation immediately after selecting the unconstrained optimum (and weighted optimums for present value budget constraints if required) while the full list of options for the segment was held in the computer's memory.

One option 'dominates' another for the same segment where the dominant option has a lower PVTTC and the same annual road agency costs in all budget-constrained years. It will have higher annual road agency costs in unconstrained years, but that is irrelevant. Choosing the dominant option over the alternative saves on PVTTC without consuming any more of the budgets in constrained years.

Another way in which one option dominates another occurs where the dominant option has the same PVTTC and a lower cost in at least one budget-constrained year without higher road agency costs in any other budget-constrained year. Choosing the dominant option saves costs in a budget-constrained year with no sacrifice of PVTTC. However, instances of this second from of dominance are unlikely to occur because, with PVTTC being a continuous variable, it would be an unlikely coincidence for two different treatment options to produce precisely the same PVTTC value. Nevertheless, the dominance test in the model, detailed in Appendix B, would find any such cases.

Formally, letting $c_t$ represent agency cost in year *t* for a given segment, option *A* dominates option *B* if either

- $PVTTC^A < PVTTC^B$ and $c_t^A \leq c_t^B$ for all years *t* that are budget-constrained, or
- $PVTTC^A = PVTTC^B$ and $c_t^A < c_t^B$ for at least one year *t* that is budget-constrained.

The numerical example in Table 6.1 uses hypothetical numbers to illustrate the dominance relationships. Four segments are shown, each with two options labelled *A* and *B*. Only the first 10 years of the analysis period are shown and budget constraints are imposed for the first five years.

- Segment 1: Option *A* dominates option *B* because they both consume $100,000 of constrained year-one funds but option A has a lower PVTTC.
- Segment 2: Option *A* dominates option *B* because they both make no demands on constrained funds in years one to five, but option A has a lower PVTTC.

- Segment 3: Option *A* dominates option *B* because, while both have the same PVTTC value, option A requires less funds in constrained year two.
- Segment 4: Neither option dominates the other because they require funds in different budget-constrained years. Both options need to be retained in the list of options to take forward to the next stage of the optimisation process.

Note that the length of the period with annual budget constraints affects the number of opportunities to eliminate dominated options. In Table 6.1, had there been 10 budget-constrained years instead of five, neither of the two options for segments one and two would be dominant because they have treatment costs in different budget-constrained years. The longer the constrained period, the smaller the number of dominated options that can be eliminated.

**Table 6.1 Numerical example illustrating the dominance concept**

*($'000)*

| Segment | 1 | | 2 | | 3 | | 4 | |
|---|---|---|---|---|---|---|---|---|
| **Option** | **A** | **B** | **A** | **B** | **A** | **B** | **A** | **B** |
| **PVTTC ($)** | 1000 | 1200 | 500 | 600 | 900 | 900 | 1000 | 1100 |
| **Treatment costs by year ($)** | | | | | | | | |
| **Year** | | | | | | | | |
| **1** | 100 | 100 | | | | | | |
| **2** | | | | | 60 | 70 | | |
| **3** | | | | | | | 140 | |
| **4** | | | | | | | | 150 |
| **5** | | | | | | | | |
| **End of budget-constrained years** | | | | | | | | |
| **6** | 120 | | | | | | | |
| **7** | | | 50 | | 100 | 80 | | |
| **8** | | | | 50 | | | | |
| **9** | | | | | | | | |
| **10** | | 150 | | | | | 100 | 100 |

Note: Columns for dominant options greyed.

The model was programmed, not only to identify the option with the lowest PVTTC for each segment in the database, but also to generate the list of non-dominated options for each segment.

- For budget constraints over the first 10 years, there were 21 059 non-dominated options, an average of 10.4 options per segment, with numbers of non-dominated options for individual segments ranging from 1 to 32.
- For budget constraints over the first 20 years, there were 272 886 non-dominated options for the entire database, an average of 134.2 options per segment, with numbers of non-dominated options for individual segments ranging from 1 to 505.

In a small number of cases, a single option dominated all others, in other words, there was just one non-dominated option. This occurred for 145 segments for the 10-year optimisation and 8 segments for the 20-year optimisation. For these segments, the solution with the lowest PVTTC had either no treatments in any of the constrained years and hence made no call on funds in the constrained years (as for segment 2 in Table 6.1), or a mandatory treatment in the first year required to meet the technical restrictions in the model and treatments in no other budget-constrained years that might be shifted to other years.

## 6.2.3 **Stage 3: Penalty method**

Even after the dramatic reduction in the number of options following removal of dominated options, there remained a huge number of possible combinations of options across all the segments. Stage 3 of the optimisation process applied the 'penalty method' to obtain a solution very close to the optimum.

Shifting a treatment from its optimal time in the unconstrained solution to another year in order to reduce the demand for funds in the optimal year, imposes a cost in the form of a higher PVTTC for the segment. The penalty method minimises the total of the cost increases for all segments by making the least costly shifts. Under the penalty method, instead of minimising PVTTC, the value minimised is PVTTC plus the sum of costs in constrained years multiplied by penalty factors.

Say funds were scarce in year one but not in year two. For segments with treatments in year one in the unconstrained optimal solution, some year-one treatments need to shift to year two in order to meet the year-one budget constraint. Under the penalty method, the decision about whether to treat a particular segment in year one or year two would be made by choosing the alternative with the lower value of $PVTTC + \lambda c_1$, where $c_1$ is the costs of the treatment in year one and $\lambda$ is the penalty factor for year one. As year-two funds are unconstrained, there is no penalty factor for year two.

The decision rule is

- leave the treatment in year one if $PVTTC_{y1} + \lambda c_1 < PVTTC_{y2}$
- shift the treatment to year two if $PVTTC_{y1} + \lambda c_1 > PVTTC_{y2}$.

where $PVTTC_{y1}$ and $PVTTC_{y2}$ are the respective PVTTCs with the treatment carried out in years one and two.

The decision rule can be rewritten as

- leave the treatment in year one if $(PVTTC_{y2} - PVTTC_{y1})/c_1 > \lambda$, that is, if the cost per dollar of year-one budget saved is above a cut-off value $\lambda$
- shift the treatment it to year two if $(PVTTC_{y2} - PVTTC_{y1})/c_1 < \lambda$, that is, if the cost per dollar of year-one budget saved is below the cut-off value $\lambda$.

As the treatment is carried out in year one in the unconstrained optimum, we know that $PVTTC_{y1} < PVTTC_{y2}$ , so the ratio $(PVTTC_{y2} - PVTTC_{y1})/c_1$ is positive. To induce the treatment to shift to year two, $\lambda$ has be raised to at least the value of the ratio. With the decision rule being applied to a number of segments together, as $\lambda$ is progressively raised, more year-one treatments shift to year two. The penalty factor, $\lambda$, would be increased to the point where the year-one budget constraint was just met. The $\lambda$ value at this point is the maximum acceptable increase in PVTTC to save a dollar of scarce year-one funds.

Table 6.2 presents a simple numerical example to illustrate application of the rule. Two options are shown for two segments. In each case, option A has a treatment in year one and option B, does not. Treatments occur in other years not shown in the table. For both segments, option A is preferred in the absence of budget constraints because it has the lower PVTTC. In the presence of a year-one funding constraint, switching from option A to option B, in both cases, increases PVTTC but saves on scarce year-one funds. In the case of segment 1, $100,000 of funds is saved at a cost of a $300,000 increase in PVTTC, a cost to benefit ratio, $(PVTTC_{y2} - PVTTC_{y1})/c_1$, of 3.0. For segment 2, $250,000 of year-one funds is saved at a cost of $50,000 in PVTTC, a cost to benefit ratio of 2.0. To reduce spending in year one, the segment 2 treatment would be delayed first because it has the lower cost to benefit ratio. Only if further savings were required, would the treatment for segment 1 be delayed. So treatments would be shifted out of year one in ascending order of cost to benefit ratio.

The last three rows of Table 6.2 show the effect of applying penalty factors of 1.5, 2.5 and 3.5. The 1.5 penalty factor still leads to option A being chosen for both segments. Increasing the penalty factor to 2.5 causes a switch to option B for segment 2 only. A further increase in the penalty factor to 3.5 causes segment 1 to switch to option B as well.

**Table 6.2 Numerical example illustrating decision rule with penalty factor**
*($'000)*

| Segment | 1 | | 2 | |
|---|---|---|---|---|
| Option | A | B | A | B |
| **PVTTC** | 1000 | 1300 | 2000 | 2500 |
| **$PVTTC_{y2} – PVTTC_{y1}$** | 300 | | 500 | |
| **Treatment cost in year 1** | 100 | 0 | 250 | 0 |
| | | | | |
| **$(PVTTC_{y2} – PVTTC_{y1}) / c_1$** | 3 | | 2 | |
| | | | | |
| **PVTTC + $\lambda c_1$ for $\lambda = 1.5$** | 1150 | 1300 | 2375 | 2500 |
| **PVTTC + $\lambda c_1$ for $\lambda = 2.5$** | 1250 | 1300 | 2625 | 2500 |
| **PVTTC + $\lambda c_1$ for $\lambda = 3.5$** | 1350 | 1300 | 2825 | 2500 |

Note: Greyed cells show the minimum *PVTTC + $\lambda$ $c_1$* values.

Faced with a choice between many segments with treatments in year one in the unconstrained optimal solution that can be shifted to year two, the penalty method prioritises shifts in ascending order of their cost (increase in PVTTC) to benefit (saving in scarce funds) ratios, minimising the overall cost of achieving the budget constraint. A $\lambda$ value of zero will shift no treatments. Setting a small $\lambda$ value identifies segments for which shifts to year two can be made for little cost in terms of a higher PVTTC. The $\lambda$ value can be progressively raised, shifting treatments out of year one until the year-one budget constraint is met.

The general form of the decision rule for multiple budget-constrained years is: where budget constraints exist for years 1 to *m*, for each individual road segment *i*, select from the list of non-dominated options, the treatment option (set of times and types) with the lowest value of

$$PVTTC_i + \sum_{t=1}^{m} \lambda_t c_{it}$$

where

- for each year *t* from 1 to *m*, a single penalty factor, $\lambda_t$, is applied to all segments in the network
- the $\lambda_t$ values are set at zero for years in which the budget constraint is non-binding, and
- for years where the constraint is binding, the penalty factor is set at a level just high enough to ensure the budget constraint for the particular year is met.

Summing the spending needs for all *n* segments in each year, the years with spending above the budget constraints ($B_t$) can be identified and their penalty factors increased to the levels at which the budget constraints are met, that is,

$$\sum_{i=1}^{n} c_{it} \leq B_t$$

for all constrained years *t*.

### 6.2.3.1 Interpretation of penalty factors

The penalty factors can be interpreted as Lagrange multipliers from which estimates of MBCRs for individual years can be obtained. In Chapter 3, optimisation of PVTTC subject to annual budget constraints was discussed in terms of minimising the Lagrangian

$$L = PVTTC(c_1, c_2, \dots, c_m, c_{m+1}, \dots) - \sum_{t=1}^{m} \lambda_t (B_t - c_t)$$

where PVTTC is assumed to be a continuous function of spending each year.

The penalty method applied to *n* segments minimises

$$\sum_{i=1}^{n}\left[PVTTC_i(c_{i1}, c_{i2}, \dots, c_{im}, c_{im+1}, \dots) + \sum_{t=1}^{m} \lambda_t c_{it}\right]$$

Since the annual budget constraints, $B_t$, are constants, they are not required in the expression to be minimised under the penalty method.

For a single segment, $min\ (PVTTC_i + \sum_{t=1}^{m} \lambda_t c_{it})$ is discontinuous as the penalty factor for a single year is changed because the model has only a limited number of options to choose from and each option will only have treatments in, at most, a few years with budget constraints. However, for a large number of segments taken together, the discontinuities in the sum of PVTTCs for all the segments, when $\lambda$ values are changed, are very small.

Provided network PVTTC as a function the penalty factors, $\sum_{i=1}^{n} PVTTC_i$, is not too discontinuous, the set of penalty factors that meets the given set of annual budget constraints can be converted to MBCRs using the formula from Chapter 3, $MBCR_t = (1+r)^t \lambda_t + 1$.

### 6.2.4 **Stage 4: Refining the solution with a genetic algorithm**

The solution from the stage 3 optimisation was the starting point for the stage 4 optimisation. In stage 4, the genetic algorithm was allowed to choose between all options for all segments for which two or more non-dominated treatment options were available. Since the genetic algorithm started with a solution already quite close to the optimum, the number of potential solutions to explore — that is, solutions that improve on the starting point — was quite limited.

Due to the discrete nature of the problem, there are likely to be some years in which funds are not fully spent following the stage 3 optimisation because the next cheapest shift into a budget-constrained year from a year with a higher $\lambda$ value exceeds the constraint. Say the constraint in year five was \$10 million, and the penalty factor optimisation method assigned treatments worth \$9.6 million to year five with the cost of the last treatment assigned at \$0.1 million. The next cheapest shift into year five, from another year in which funds were scarcer (a higher $\lambda$ value), involves a single treatment costing \$0.5 million. This shift would not be made because it would increase year-five spending to \$10.1 million, which exceeds the constraint. Allowing the genetic algorithm to finesse the solution by testing solutions with different options for each segment, it could be found that the network PVTTC is reduced by shifting the \$0.1 million treatment from year five to another year in order to make room for the \$0.5 million treatment in year five. The \$10 million constraint in year five would then be fully utilised.

## 6.3 Implementing the methodology

For each segment, the stage 1 optimisation passed to stage 2 a list of all options that were technically feasible (the 'all-options list'), each with present values of total, agency and user costs and treatment costs. Stage 2 involved reducing this to a list of non-dominated options. The list of non-dominated options was built up starting with the first option in the all-options list as the seed. Beginning with the second option in the all-options list, each option in the all-options list was in turn compared with every option in the non-dominated list. If it dominated one or more members of the current non-dominated list, the dominated members were removed from the list. If was not itself dominated by any members of the non-dominated list, it was added to the list. The process was complete after the last option in the all-options list had been compared with all options in the non-dominated list. Further detail is provided in Appendix B, Section B.2.

Having obtained a list of non-dominated options for each segment in the database from the model implemented in Mathematica (stages 1 and 2), the remaining two stages (3 and 4) were undertaken using Excel linked to Evolver genetic algorithm software.

In stage 3, a penalty factor or $\lambda$ value had to be set for each of the *m* budget-constrained years, such that that spending in each year was within the budget constraint for that year. For each individual segment *i*, the spreadsheet was set up to select, from the list of non-dominated options, the option that minimises $PVTTC_i + \sum_{t=1}^{m} \lambda_t c_{it}$ for the segment by itself. Each segment then had a single selected option. The sum of spending over all segments in each year gave total spending for the year.

If all penalty factors were set at zero, the spreadsheet provided the unconstrained optimal solution with the option with the lowest PVTTC value selected for each segment. If the penalty factor for a given year was increased, the imposition of an additional penalty on costs for that year caused treatments for some segments to move out of that year as options were selected with treatments in other years. Although the model was discrete, it behaved almost as if it were continuous. Increasing the penalty factors for particular years is like pressing down on a viscous substance — the mass springs up on either side. However, unlike a physical substance, the total volume does not remain constant. Small reductions in annual spending constraints could save agency costs at the expense of users. However, large reductions in the early years increased overall agency costs due to the principle of 'a stitch in time saves nine'.

It proved impossible to set the penalty factors manually. Adjusting one penalty factor to cut spending in one year to meet a constraint caused increases in spending in other years, breaching the constraints in those years and requiring further adjustments. This is known as the 'waterbed effect'. As the problem was non-smooth, a gradient (steepest descent) method could not be used.

An Excel macro, which adjusted $\lambda$'s downward by a specified proportion for all years where constraints were non-binding, and upward by the same proportion for all years for which spending was above the constraint, was able to iterate towards an approximate solution. The adjustment proportions were decreased as spending in each year approached the constraints.

The Evolver genetic algorithm software was employed to complete adjustment of the 10 or 20 annual penalty factors. The objective was initially set to minimise the maximum extent to which constraints were exceeded, that is, to minimise $\max_t\left(\sum_{i=1}^{2034} c_{it} - B_t, 0\right)$, where $c_{it}$ is the cost for segment *i* in year *t* and $B_t$ the budget for year *t*. This identified the year with the greatest constraint violation and focused on reducing it. Once all constraints were met, the objective was changed to minimising the sum of $\lambda$'s subject to the budget constraints. It was not necessary to minimise $\sum_{i=1}^{n}(PVTTC_i + \sum_{t=1}^{m} \lambda_t c_{it})$ because the spreadsheet automatically selected the option that minimises $PVTTC_i + \sum_{t=1}^{m} \lambda_t c_{it}$ for each segment, given the penalty factors.

It was important to take the optimisation process to the point where no further possible improvements could be made in terms of reducing $\lambda$ values. The budget constraints could be met by solutions that have higher $\lambda$ values than necessary. Failure to ensure the constraints are met with the lowest possible $\lambda$ values leads to exaggerated MBCR estimates. This is why the step in which the sum of $\lambda$ values was minimised was essential.

Stage 3 could have been undertaken within Mathematica by developing an algorithm along the lines of the Excel macro employed, however, Excel was necessary for Stage 4 undertaken using Evolver because Evolver only operates within Excel. As the number of variables was very large, one for each segment, it was necessary to use Evolver Industrial Edition, which can handle an unlimited number of adjustable variables.

## 6.4 Case study results: annual budget constraints for 10 years

Table 6.3 shows results, in descending order of tighter uniform annual budget constraints, for the first 10 years starting with the unconstrained result.

**Table 6.3 Model results for 10-year uniform annual budget constraints**

*($ millions)*

| Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Constraints** | nil | | | | | | | | | | nil | |
| **Spending** | 185.8 | 51.9 | 69.8 | 80.6 | 121.3 | 102.8 | 79.7 | 53.1 | 85.2 | 73.2 | 82.3 | 93.4 |
| **Lambdas** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| **MBCRs** | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| **Constraints** | 100 | | | | | | | | | | nil | |
| **Spending 3** | 99.6 | 99.8 | 99.3 | 98.4 | 98.8 | 97.4 | 98.4 | 88.3 | 66.4 | 81.9 | 95.6 | 82.5 |
| **Spending 4** | 100.0 | 100.0 | 99.9 | 97.3 | 99.9 | 99.5 | 99.9 | 83.7 | 66.5 | 81.8 | 96.1 | 81.4 |
| **Lambdas** | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| **MBCRs** | 1.2 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| **Constraints** | 80 | | | | | | | | | | nil | |
| **Spending 3** | 80.0 | 80.0 | 79.9 | 79.1 | 79.7 | 75.2 | 79.9 | 79.8 | 78.6 | 79.9 | 194.0 | 117.1 |
| **Spending 4** | 80.0 | 80.0 | 80.0 | 80.0 | 80.0 | 80.0 | 80.0 | 79.7 | 79.4 | 79.9 | 187.4 | 117.0 |
| **Lambdas** | 0.5 | 0.2 | 0.2 | 0.1 | 0.1 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| **MBCRs** | 1.5 | 1.3 | 1.2 | 1.2 | 1.1 | 1.1 | 1.1 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| **Constraints** | 60 | | | | | | | | | | nil | |
| **Spending 3** | 60.0 | 59.9 | 59.7 | 60.0 | 60.0 | 56.0 | 59.6 | 59.9 | 59.6 | 55.4 | 477.1 | 109.5 |
| **Spending 4** | 60.0 | 60.0 | 60.0 | 60.0 | 60.0 | 60.0 | 60.0 | 60.0 | 60.0 | 60.0 | 470.3 | 109.5 |
| **Lambdas** | 1.4 | 1.0 | 0.7 | 0.5 | 0.4 | 0.3 | 0.2 | 0.1 | 0.1 | 0.0 | 0.0 | 0.0 |
| **MBCRs** | 2.4 | 2.1 | 1.8 | 1.6 | 1.5 | 1.4 | 1.3 | 1.2 | 1.1 | 1.1 | 1.0 | 1.0 |
| **Constraints** | 50 | | | | | | | | | | nil | |
| **Spending 3** | 50.0 | 50.0 | 49.9 | 49.8 | 49.9 | 47.6 | 46.9 | 49.1 | 48.0 | 49.7 | 796.6 | 73.7 |
| **Spending 4** | 50.0 | 50.0 | 50.0 | 50.0 | 50.0 | 50.0 | 50.0 | 49.5 | 50.0 | 50.0 | 788.0 | 72.9 |
| **Lambdas** | 2.9 | 2.3 | 1.8 | 1.4 | 1.1 | 0.8 | 0.5 | 0.3 | 0.3 | 0.1 | 0.0 | 0.0 |
| **MBCRs** | 4.0 | 3.4 | 3.0 | 2.6 | 2.3 | 2.0 | 1.7 | 1.5 | 1.4 | 1.2 | 1.0 | 1.0 |
| **Constraints** | 40 | | | | | | | | | | nil | |
| **Spending 3** | 39.8 | 40.0 | 39.8 | 40.0 | 40.0 | 39.9 | 34.8 | 40.0 | 39.9 | 39.7 | 1382.9 | 135.9 |
| **Spending 4** | 40.0 | 40.0 | 40.0 | 40.0 | 40.0 | 40.0 | 40.0 | 40.0 | 40.0 | 39.9 | 1379.5 | 133.9 |
| **Lambdas** | 7.7 | 6.4 | 5.2 | 4.3 | 3.9 | 2.8 | 1.8 | 1.2 | 0.9 | 0.6 | 0.0 | 0.0 |
| **MBCRs** | 9.0 | 7.9 | 6.9 | 6.0 | 5.7 | 4.5 | 3.4 | 2.7 | 2.3 | 2.0 | 1.0 | 1.0 |
| **Constraints** | 35.3 | | | | | | | | | | nil | |
| **Spending 3** | 35.3 | 35.2 | 35.3 | 35.1 | 33.5 | 35.3 | 33.1 | 34.9 | 35.1 | 35.0 | 1619.5 | 288.3 |
| **Spending 4** | 35.3 | 35.3 | 35.3 | 35.3 | 35.3 | 35.3 | 35.3 | 35.2 | 35.3 | 35.3 | 1615.9 | 287.9 |
| **Lambdas** | 24.6 | 20.8 | 17.5 | 14.6 | 12.3 | 9.0 | 6.8 | 4.0 | 3.1 | 2.1 | 0.0 | 0.0 |
| **MBCRs** | 26.5 | 23.4 | 20.7 | 18.1 | 16.0 | 12.4 | 10.0 | 6.5 | 5.4 | 4.2 | 1.0 | 1.0 |

Notes: Spending 3 is results from the stage 3 optimisation using penalty factors. Spending 4 is results after the stage 4 optimisation letting the genetic algorithm select individual options. Although the budget constraints were imposed for years 1 to 10 only, years 11 and 12 are included the table to show how spending was pushed out into the first few unconstrained years.

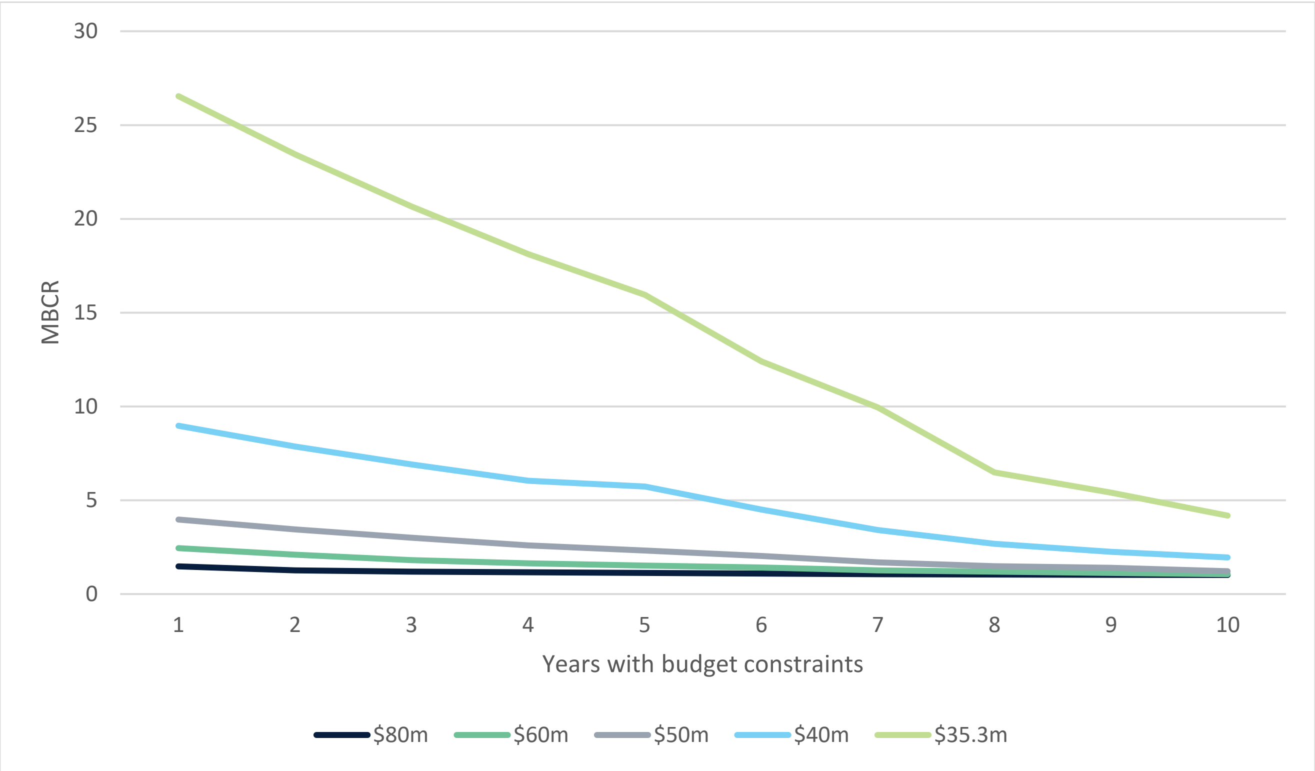
Spending levels for years 11 and 12 are included to show that imposing the constraints pushes out maintenance spending into the period immediately following the constrained period. The 'spending 3' rows are the results obtained from the stage 3 optimisation using only the penalty method. The 'spending 4' rows are results after the stage 4 optimisation in which the stage 3 solution was refined by having the genetic algorithm select individual options. Also, shown are the annual MBCR values obtained from the penalty factors.

The first set of constraints, at $100 million, is above the 10-year average annual spending level of $90.3 million for the unconstrained scenario. As the constraints were progressively tightened, the penalty factors had to be raised to induce more expensive shifts in treatment times. The lowest level of 10-year uniform annual budget constraint obtainable was $35.3 million, found by having the model adjust the penalty factors to minimise the maximum of the annual spending levels across the 10 years. This led to MBCRs above 20 in the early years and a very large amount of spending pushed out into years 11 and 12. Indeed, the scenarios with budget constraints of $60 million per annum and below might be considered unrealistic because of the large increases in spending required in years 11 and 12 to catch up the backlog.

The results for the $100 million constraints show that shifting $86 million of unconstrained first-year spending into subsequent years cost little. The MBCR for the first year with the $100 million constraint was only 1.2. The increase in PVTTC was just $4.5 million. The reason it was so low is that the model had shifted the year-one treatments that were the cheapest to delay. Tightening of budget constraints at modest levels leads to gentle rises in MBCRs but extremely steep rises are required as constraints approach the limit. This relationship was observed in Chapter 5 for present value budget constraints.

Also evident is that, due to the higher demands for spending in early years (due to the maintenance backlog), annual budget constraints that are uniform across years require declining penalty factors starting from the highest in year one when demand for spending is greatest. Figure 6.1 plots the annual MBCR values under each set of budget constraints below $100 million showing the higher MBCRs associated with tighter constraints, and how they are highest in the first year and decline throughout the constrained period.

**Figure 6.1 Annual MBCR values for 10-year uniform annual budget constraints**




It is understood that actual spending on the case study network around the time the data was compiled was around $15 million to $20 million. It proved impossible to force the model to meet spending constraints in all of the first 10 years around this level. It was possible to force spending down to $20 million for the first five years with no constraints thereafter. Under the latter case, MBCRs for the five years averaged 10.0 with a backlog of spending in year six, the first unconstrained year, of $703 million. If only the first four years were constrained to $20 million, the average MBCR for the first four years averaged 4.1 and the backlog of spending on year five was $519 million. This illustrates how relaxing pressure on budgets in later years, lowers

MBCRs in earlier years. The MBCR values obtained from constraining a small number of years followed by a large backlog in the first unconstrained year are misleading because they are premised on an unrealistic assumption of unlimited financial and physical capacity to undertake work in the first unconstrained year.

Table 6.4 shows the changes in present values caused by annual budget constraints compared with the unconstrained optimum. The increases in PVTTC show that loose budget constraints come at a small cost to society but then increase rapidly as the constraints approach the tightest possible level. It is interesting that PVAC is lower for the $80 million and $60 million per annum constraints but then increases steeply as the large amounts of catch-up spending after year 10 outweigh the cost savings to the road agency in the constrained years up to year 10. The catch-up spending after year 10 was so great as to make users better off in present value terms with the $35.3 million constraints. To ascertain why this occurs, model results for individual segments with lower PVUC values with tight constraints were examined in detail. It was found that less expensive treatments undertaken during the 10 constrained years under less constrained scenarios were being sacrificed at a cost of bringing forward in time more expensive treatments in the later unconstrained years. These more expensive treatments have much greater impacts on improving road roughness, which determines user costs. However, the required amounts of catch-up spending after year 10 were impractically high.

Table 6.3 compares the results of the stage 3 and stage 4 optimisations for annual spending. The fourth-stage optimisation allowed Evolver to improve on the solution by selecting individual options for each segment. As can be seen in Table 6.3, in the stage 3 optimisation, lumpiness in the treatment options available prevented the model from fully utilising the budgets for some years. The stage 4 optimisation improved the utilisation of available budgets and the PVTTC values.

Table 6.4 shows the improvements in PVTTC brought about by the stage 4 optimisation process with PVUC normalised to zero at the unconstrained option. The improvements were minor. Their size increased as the constraints tightened, where the model, in stage 3 using the penalty method, had less flexibility to shift treatments between years. The PVTTC gain was achieved by saving user costs at the expense of agency costs, because fully utilising constraints brings forward spending. The percentage of the 2034 of segments with treatments changed in stage 4 ranged from 2% to 6% depending on the set of constraints.

**Table 6.4 Impacts on present values of 10-year uniform annual budget constraints and of stage 4 optimisation**

*($ millions)*

| Constraints | Optimisation stage | PVTTC | PVAC | PVUC |
|---|---|---|---|---|
| **Nil** | | 1902.48 | 1902.48 | 0 |
| **100** | Stage 3 | 1906.97 | 1902.62 | 4.35 |
| | Stage 4 | 1906.94 | 1902.97 | 3.96 |
| | Improvement | –0.03 | 0.35 | –0.38 |
| **80** | Stage 3 | 1921.02 | 1883.63 | 37.38 |
| | Stage 4 | 1920.71 | 1885.01 | 35.70 |
| | Improvement | –0.30 | 1.38 | –1.68 |
| **60** | Stage 3 | 1974.19 | 1882.22 | 91.97 |
| | Stage 4 | 1973.55 | 1883.09 | 90.45 |
| | Improvement | –0.64 | 0.88 | –1.52 |
| **50** | Stage 3 | 2048.68 | 1912.81 | 135.87 |
| | Stage 4 | 2047.19 | 1913.78 | 133.41 |
| | Improvement | –1.49 | 0.97 | –2.46 |
| **40** | Stage 3 | 2258.35 | 2180.18 | 78.17 |
| | Stage 4 | 2255.39 | 2180.44 | 74.95 |
| | Improvement | –2.96 | 0.26 | –3.22 |
| **35.3** | Stage 3 | 2514.63 | 2564.76 | –50.13 |
| | Stage 4 | 2510.59 | 2565.22 | –54.63 |
| | Improvement | –4.04 | 0.46 | –4.50 |

Notes: PVUC at the unconstrained optimum has been normalised to zero. Hence, PVUC values in the constrained scenarios are the excess over the PVUC value at the unconstrained optimum. The 'improvement' is the stage 4 present value result minus the stage 3 present value result.

Table 6.5 again shows the present values with PVUC normalised to zero at the unconstrained optimum, but only after the stage 4 optimisation. It highlights the changes in the present values as the constraints were tightened and calculates incremental BCRs for annual budget constraints as defined in Chapter 3, Section 3.4.5 as

$$IBCR = -\frac{\Delta PVUC + \Delta PVAC - \Delta PVB}{\Delta PVB} = -\frac{\Delta PVTTC}{\Delta PVB} + 1$$

where PVB is the present value of the constrained annual budgets, in this case over 10 years. The right-most column shows the simple averages of the 10 MBCRs in Table 6.3 for the start and end constraint levels between each IBCR. With the exception of the comparison with the unconstrained scenario, the IBCRs between each pair of constraints lie between the average MBCRs for the constraints, which is to be expected. The reason for the exception in the case of the IBCR comparing the $100 million constraint with the unconstrained scenario, is that, for seven of the 10 years, unconstrained optimal spending was under the $100 million level, leaving a great deal of capacity to shift spending between years at little cost.[30]

----------

[30] Footnote 14 in Section 3.4.5 discussed and recommended against an alternative definition of the IBCR for increases in annual budget constraints with ΔPVAC in the denominator instead of ΔPVB. Using the alternative definition, the IBCRs for the six rows in Table 6.5 are –8.0, 1.8, 28.6, –1.4, 0.2, and 0.3 respectively. The high value of 28.6 occurred at the point where PVAC reached a minimum and then began to rise as budget constraints were tightened, causing ΔPVAC to be small at –$2 million.

**Table 6.5 Present values and BCRs for 10-year uniform annual budget constraints**

*($ millions)*

| | Constraint | PVTTC | PVAC | PVUC | PVB | IBCR | Avg MBCR |
|---|---|---|---|---|---|---|---|
| | – | 1902 | 1902 | 0 | 747 | | 1.0 |
| Δ | **–** | **4** | **0** | **4** | **–13** | **0.7** | |
| | 100 | 1907 | 1903 | 4 | 760 | | 1.0 |
| Δ | **–20** | **14** | **–18** | **32** | **–112** | **1.1** | |
| | 80 | 1921 | 1885 | 36 | 648 | | 1.1 |
| Δ | **–20** | **53** | **–2** | **55** | **–162** | **1.3** | |
| | 60 | 1974 | 1883 | 90 | 487 | | 1.6 |
| Δ | **–10** | **74** | **31** | **43** | **–81** | **1.9** | |
| | 50 | 2047 | 1914 | 133 | 405 | | 2.3 |
| Δ | **–10** | **208** | **267** | **–58** | **–81** | **3.6** | |
| | 40 | 2255 | 2180 | 75 | 324 | | 5.0 |
| Δ | **–4.7** | **255** | **385** | **–130** | **–38** | **7.7** | |
| | 35.3 | 2511 | 2565 | –55 | 286 | | 14.3 |

Notes: The values of PVTTC were calculated with PVUC set to zero at the unconstrained optimum as in Table 6.4.
PVB is the present value of road agency spending over the 10 constrained years after the last stage of optimisation.
The rows headed Δ show differences between the rows just above and just below.
The IBCRs are comparisons between the rows just above and just below, not comparisons with the first row showing unconstrained spending.
The average MBCR is the simple average of the MBCRs for the 10 budget-constrained years in Table 6.3.

## 6.5 Case study results: annual budget constraints for 20 years

Table 6.6 shows model results for three sets of constant annual budget constraints imposed over a period of 20 years in the same format as for Table 6.3. The $100 million constraint scenario for 20 years is not shown because it is identical to the $100 million 10-year scenario. Under the 10-year $100 million scenario, optimal spending in all years from 11 to 20 was below the $100 million amount so no changes were needed to extend the constraints out for a further 10 years.

Uniform constraints over the 20-year period at $80 million per annum could be considered ‘sustainable’ and gave rise to an MBCR of 1.5 in year one and no jump in spending in year 21. A ’sustainable’ level of spending could be defined as one where there is no jump in optimal spending just following the constrained period, or the jump is not so large that it cannot be caught up by continued spending at the sustainable level in in subsequent years. The lowest achievable uniform constraint over 20 years was $48.4 million, which lead to an MBCR of 21.4 in year one and a $0.9 billion demand for spending in year 21. A small easing of the constraint to $50 million lowered the year-one MBCR to 11.0 and but still left a $0.8 billion backlog of spending in year 21.

## Table 6.6 Model results for 20-year uniform annual budget constraints

*($ millions)*

| Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Unconstrained spending** | 185.8 | 51.9 | 69.8 | 80.6 | 121.3 | 102.8 | 79.7 | 53.1 | 85.2 | 73.2 | 82.3 | 93.4 | 62.7 | 65.8 | 35.1 | 37.0 | 76.0 | 46.8 | 48.6 | 54.0 | 49.0 | 37.1 |
| **Constraints** | 80 | | | | | | | | | | | | | | | | | | | | nil | |
| **Spending 3** | 80.0 | 80.0 | 79.8 | 79.4 | 79.8 | 79.5 | 78.1 | 79.0 | 79.2 | 79.6 | 77.0 | 77.9 | 79.0 | 70.8 | 79.7 | 61.4 | 78.5 | 74.9 | 78.8 | 72.7 | 61.7 | 44.4 |
| **Spending 4** | 80.0 | 80.0 | 80.0 | 80.0 | 80.0 | 80.0 | 79.9 | 80.0 | 80.0 | 80.0 | 79.7 | 79.7 | 79.8 | 79.8 | 69.7 | 60.8 | 79.8 | 75.3 | 76.9 | 65.3 | 57.6 | 44.0 |
| **Lambdas** | 0.5 | 0.3 | 0.2 | 0.2 | 0.1 | 0.1 | 0.1 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| **MBCRs** | 1.5 | 1.3 | 1.2 | 1.2 | 1.2 | 1.1 | 1.1 | 1.1 | 1.1 | 1.1 | 1.1 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| **Constraints** | 60 | | | | | | | | | | | | | | | | | | | | nil | |
| **Spending 3** | 59.7 | 60.0 | 60.0 | 56.7 | 59.9 | 60.0 | 59.1 | 58.4 | 57.3 | 59.8 | 58.8 | 44.3 | 57.5 | 55.4 | 57.1 | 58.8 | 59.9 | 59.5 | 59.1 | 52.9 | 524.7 | 81.5 |
| **Spending 4** | 60.0 | 60.0 | 60.0 | 60.0 | 60.0 | 60.0 | 60.0 | 60.0 | 60.0 | 60.0 | 60.0 | 60.0 | 60.0 | 59.9 | 59.9 | 59.9 | 60.0 | 60.0 | 60.0 | 60.0 | 481.4 | 67.9 |
| **Lambdas** | 2.4 | 1.9 | 1.5 | 1.2 | 1.0 | 0.8 | 0.6 | 0.5 | 0.4 | 0.4 | 0.3 | 0.3 | 0.2 | 0.2 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.0 | 0.0 | 0.0 |
| **MBCRs** | 3.5 | 3.1 | 2.7 | 2.4 | 2.2 | 2.0 | 1.8 | 1.7 | 1.6 | 1.6 | 1.5 | 1.4 | 1.4 | 1.3 | 1.2 | 1.2 | 1.2 | 1.2 | 1.1 | 1.1 | 1.0 | 1.0 |
| **Constraints** | 50 | | | | | | | | | | | | | | | | | | | | nil | |
| **Spending 3** | 49.9 | 50.0 | 49.7 | 49.5 | 49.6 | 49.9 | 49.6 | 49.9 | 48.4 | 49.6 | 49.4 | 49.4 | 49.8 | 48.8 | 47.6 | 49.5 | 44.9 | 47.5 | 35.1 | 49.3 | 865.2 | 56.5 |
| **Spending 4** | 50.0 | 50.0 | 50.0 | 50.0 | 50.0 | 50.0 | 50.0 | 50.0 | 50.0 | 50.0 | 50.0 | 50.0 | 50.0 | 50.0 | 50.0 | 50.0 | 50.0 | 49.9 | 50.0 | 50.0 | 830.3 | 56.8 |
| **Lambdas** | 9.6 | 8.1 | 6.7 | 5.7 | 4.8 | 3.8 | 2.7 | 2.1 | 1.7 | 1.4 | 1.2 | 1.0 | 0.8 | 0.6 | 0.5 | 0.4 | 0.3 | 0.2 | 0.2 | 0.1 | 0.0 | 0.0 |
| **MBCRs** | 11.0 | 9.7 | 8.6 | 7.6 | 6.8 | 5.8 | 4.6 | 3.9 | 3.4 | 3.1 | 2.8 | 2.5 | 2.3 | 2.0 | 1.8 | 1.7 | 1.6 | 1.5 | 1.4 | 1.2 | 1.0 | 1.0 |
| **Constraints** | 48.4 | | | | | | | | | | | | | | | | | | | | nil | |
| **Spending 3** | 48.3 | 48.4 | 48.2 | 48.3 | 47.6 | 45.9 | 43.7 | 48.3 | 48.1 | 45.8 | 48.3 | 48.2 | 48.4 | 47.0 | 47.4 | 35.9 | 48.0 | 48.4 | 47.7 | 26.7 | 965.3 | 81.0 |
| **Spending 4** | 48.4 | 48.4 | 48.4 | 48.4 | 48.4 | 48.4 | 48.4 | 48.4 | 48.4 | 48.4 | 48.4 | 48.4 | 48.4 | 48.4 | 48.4 | 48.3 | 48.4 | 48.4 | 48.4 | 48.4 | 925.1 | 74.0 |
| **Lambdas** | 19.7 | 16.8 | 14.3 | 12.2 | 10.6 | 8.8 | 5.9 | 4.1 | 3.4 | 2.8 | 2.3 | 1.9 | 1.6 | 1.2 | 0.9 | 0.7 | 0.6 | 0.5 | 0.5 | 0.3 | 0.0 | 0.0 |
| **MBCRs** | 21.4 | 19.2 | 17.1 | 15.3 | 13.9 | 12.2 | 8.7 | 6.7 | 5.8 | 5.1 | 4.5 | 4.0 | 3.6 | 3.1 | 2.6 | 2.3 | 2.1 | 2.0 | 2.0 | 1.6 | 1.0 | 1.0 |

Notes: Spending 3 is results from the stage 3 optimisation using penalty factors. Spending 4 is results after the stage 4 optimisation letting the genetic algorithm select individual options. Although the budget constraints were imposed for years 1 to 20 only, years 21 and 22 were included the table to show how spending was pushed out into the first few unconstrained years.

## Table 6.7 Model results for 20-year rising annual budget constraints

*($ millions)*

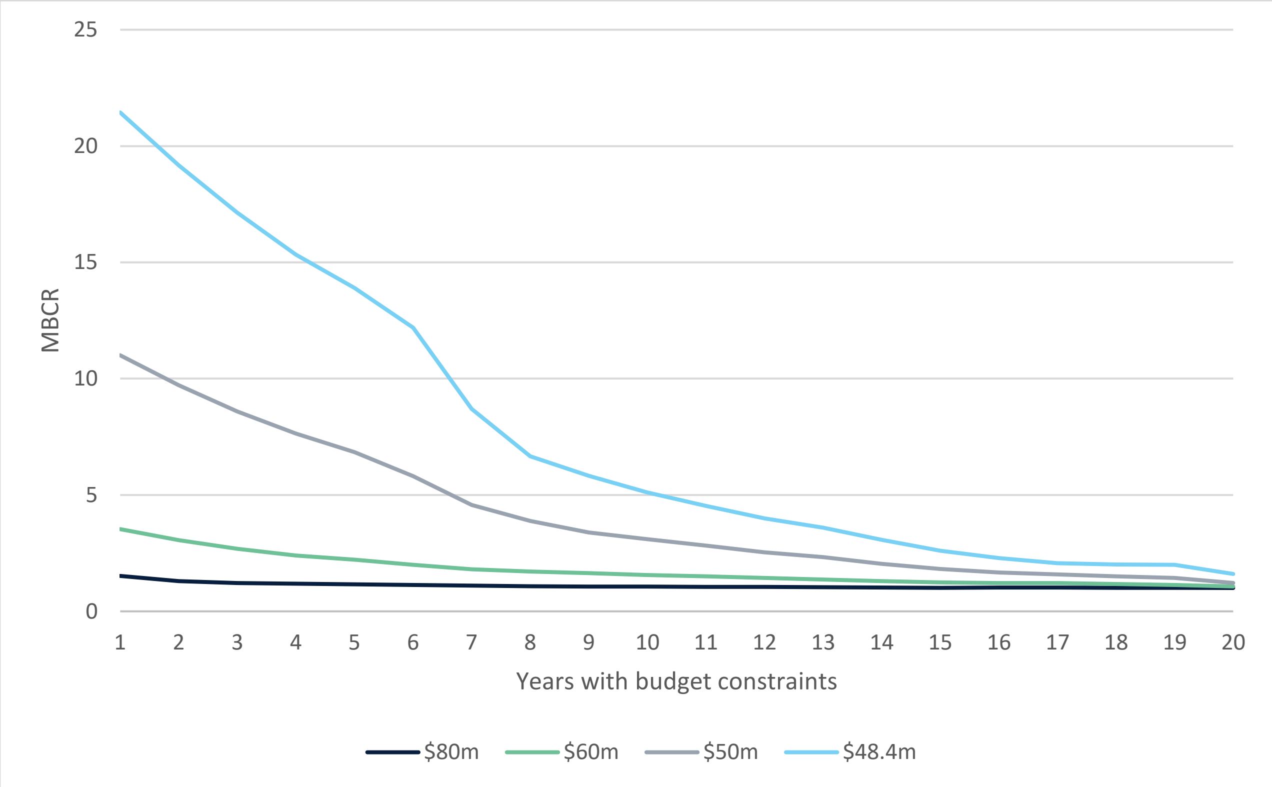
| Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Constraints** | 30 | 40 | 50 | 60 | 70 | 80 | 90 | | | | | | | | | | | | | | nil | |
| **Spending 3** | 30.0 | 39.3 | 49.8 | 60.0 | 68.8 | 79.5 | 90.0 | 88.8 | 89.2 | 83.2 | 88.4 | 65.5 | 81.9 | 89.5 | 83.3 | 89.9 | 81.8 | 67.7 | 89.9 | 80.7 | 244.2 | 36.8 |
| **Spending 4** | 30.0 | 40.0 | 50.0 | 60.0 | 70.0 | 80.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 89.9 | 89.7 | 88.5 | 84.5 | 89.3 | 88.9 | 74.8 | 89.2 | 77.3 | 194.4 | 28.0 |
| **Lambdas** | 3.8 | 2.9 | 2.3 | 1.8 | 1.3 | 1.0 | 0.7 | 0.5 | 0.4 | 0.3 | 0.3 | 0.2 | 0.2 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 |
| **MBCRs** | 5.0 | 4.2 | 3.6 | 3.0 | 2.6 | 2.2 | 1.9 | 1.7 | 1.6 | 1.5 | 1.4 | 1.3 | 1.3 | 1.2 | 1.2 | 1.1 | 1.1 | 1.1 | 1.1 | 1.0 | 1.0 | 1.0 |
| **Constraints** | 25 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | | | | | | | | | | | | | nil | |
| **Spending 3** | 24.9 | 30.0 | 40.0 | 49.9 | 52.2 | 69.9 | 79.8 | 88.4 | 90.0 | 89.7 | 88.3 | 90.0 | 72.2 | 87.4 | 61.4 | 87.7 | 84.6 | 79.5 | 84.6 | 72.9 | 532.7 | 43.4 |
| **Spending 4** | 25.0 | 30.0 | 40.0 | 50.0 | 60.0 | 70.0 | 80.0 | 90.0 | 90.0 | 90.0 | 89.9 | 90.0 | 89.9 | 90.0 | 89.8 | 89.5 | 77.8 | 89.6 | 89.8 | 90.0 | 441.2 | 35.8 |
| **Lambdas** | 19.5 | 13.7 | 10.6 | 8.5 | 7.0 | 5.4 | 3.7 | 2.3 | 1.7 | 1.3 | 1.0 | 0.8 | 0.6 | 0.4 | 0.3 | 0.2 | 0.2 | 0.1 | 0.1 | 0.1 | 0.0 | 0.0 |
| **MBCRs** | 21.3 | 15.8 | 12.9 | 10.9 | 9.5 | 7.8 | 5.8 | 4.2 | 3.4 | 2.9 | 2.6 | 2.3 | 2.0 | 1.7 | 1.6 | 1.5 | 1.4 | 1.3 | 1.2 | 1.1 | 1.0 | 1.0 |
| **Constraints** | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | | | | | | | | | | | | | nil | |
| **Spending 3** | 20.0 | 29.8 | 39.4 | 49.8 | 54.7 | 69.8 | 78.3 | 88.2 | 89.9 | 90.0 | 85.2 | 89.7 | 89.7 | 89.9 | 74.6 | 89.6 | 86.6 | 89.6 | 86.3 | 89.3 | 664.0 | 55.9 |
| **Spending 4** | 20.0 | 30.0 | 40.0 | 50.0 | 60.0 | 70.0 | 80.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 89.9 | 90.0 | 631.3 | 53.1 |
| **Lambdas** | 66.2 | 38.8 | 28.2 | 23.0 | 19.5 | 15.5 | 11.1 | 6.5 | 4.5 | 3.4 | 3.0 | 2.3 | 1.6 | 1.2 | 0.8 | 0.6 | 0.5 | 0.4 | 0.3 | 0.2 | 0.0 | 0.0 |
| **MBCRs** | 69.8 | 42.9 | 32.7 | 27.9 | 24.7 | 20.6 | 15.6 | 9.9 | 7.4 | 6.1 | 5.6 | 4.7 | 3.7 | 3.0 | 2.5 | 2.1 | 1.9 | 1.8 | 1.6 | 1.3 | 1.0 | 1.0 |

Notes: Spending 3 is results from the stage 3 optimisation using penalty factors. Spending 4 is results after the stage 4 optimisation letting the genetic algorithm select individual options. Although the budget constraints were imposed for years 1 to 20 only, years 21 and 22 were included the table to show how spending was pushed out into the first few unconstrained years.

Extending the constrained period from 10 to 20 years caused the MBCRs in earlier years to be higher. To illustrate, for uniform annual constraints of $60 million and $50 million, the first-year MBCRs were respectively 2.4 and 4.0 for 10 constrained years and 3.5 and 11.0 for 20 constrained years.

As with the 10 years of constraints, MBCRs for the 20 years declined over the constrained period, illustrated in Figure 6.2.

**Figure 6.2 Annual MBCR values for 20-year uniform annual budget constraints**




It was noted above that actual spending on the case study network around the time of the data was around $15 million to $20 million but spending on these levels could only be sustained for a few years given the technical constraints in the modelling. Where budget constraints are tight to the point that they cannot possibly be sustained for more than a few years, the most realistic approach may be to have the annual budget constraints gradually arising until they reach a sustainable level. Table 6.7 shows model results for rising constraints starting at $30, $25 and $20 million and increasing in steps to a constant level of $90 million for the remainder of the budget-constrained years.

The increasing constraint scenarios shown gave rise to large spikes in spending in year 21, but this is far in the future and spending levels in year 22, and for subsequent years (not shown in the table), are well under the $90 million level. With funds so scarce in the early years, a small tightening of constraints in the early years requires a large increase in the penalty factors and hence the MBCRs. The second two rising constraint scenarios are identical for all years except year one, for which the constraint was tightened from $25 to $20 million. The year one MBCR increased enormously from 21 to 70 as large penalty factors needed to be imposed to force highly beneficial treatments out of the early years. Figure 6.3 illustrates the MBCRs over time with the increasing constraints.

**Figure 6.3 Annual MBCR values for 20-year rising annual budget constraints**

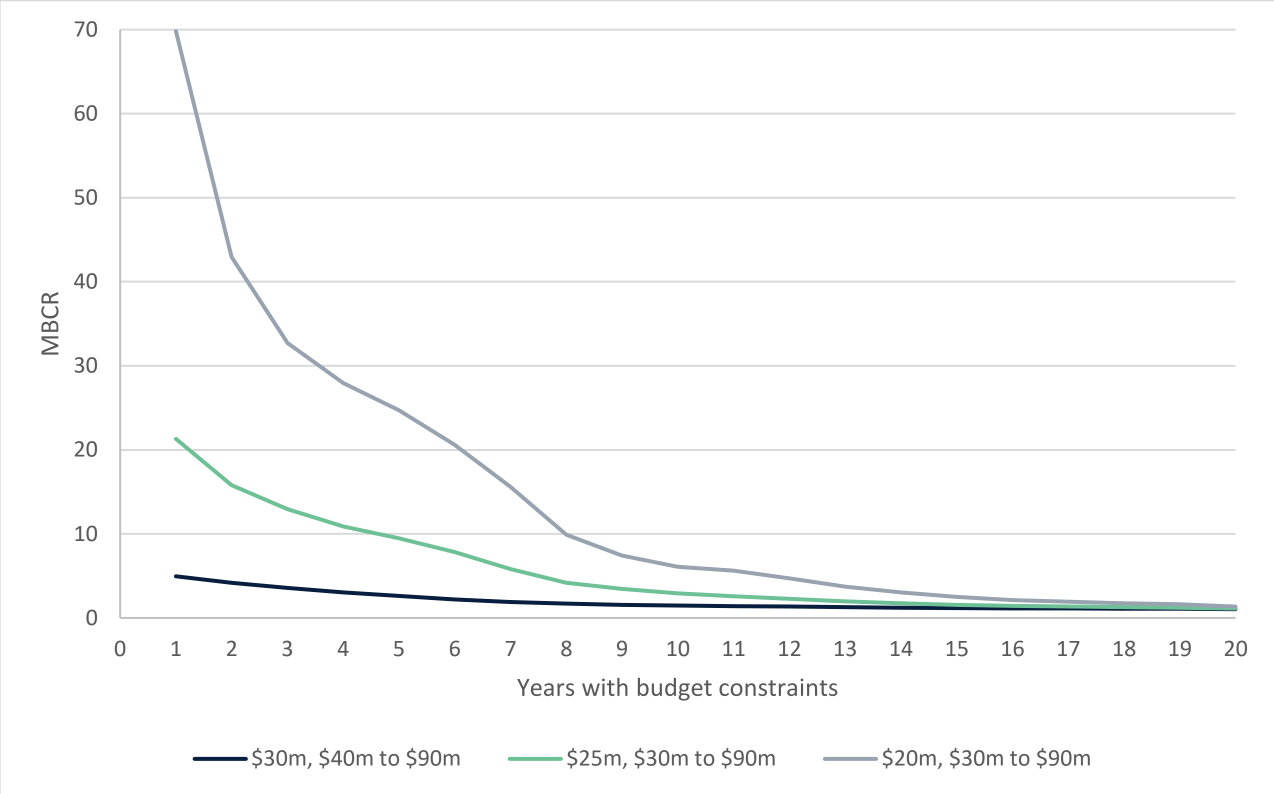



Table 6.8 shows the changes in present values compared with the optimum for the six 20-year constraint scenarios in the previous two tables.

The year-by-year results for the 20-year constraints in Tables 6.6 and 6.7 further illustrate the effect of the fourth and final stage optimisation in shifting lumpy treatments that can be moved at little cost between years to make use of unspent funds in some years following stage 3 of the optimisation process. Table 6.8 shows the impact of the fourth stage on present values. Just as Table 6.4 showed for the 10-year constraints, the improvements are small but increase as the constraints become tighter. Percentages of the 2034 segments with changed options ranged from 5% to 7% for the uniform constraints and 6% to 11% for the rising constraints.

**Table 6.8 Impacts on present values of 20-year uniform annual budget constraints and of stage 4 optimisation**

*($ millions)*

| Constraints | Optimisation stage | PVTTC | PVAC | PVUC |
|---|---|---|---|---|
| **Nil** | | 1902.5 | 1902.5 | 0 |
| **80** | Stage 3 | 1922.7 | 1868.9 | 53.8 |
| | Stage 4 | 1922.3 | 1871.0 | 51.3 |
| | Improvement | –0.3 | 2.2 | –2.5 |
| | | | | |
| **60** | Stage 3 | 2037.9 | 1806.9 | 230.9 |
| | Stage 4 | 2033.2 | 1812.0 | 221.2 |
| | Improvement | –4.6 | 5.1 | -9.7 |
| **50** | Stage 3 | 2277.8 | 1820.3 | 457.6 |
| | Stage 4 | 2272.0 | 1822.7 | 449.4 |
| | Improvement | –5.8 | 2.4 | –8.2 |
| **48.4** | Stage 3 | 2420.7 | 1853.0 | 567.7 |
| | Stage 4 | 2403.3 | 1856.7 | 546.6 |
| | Improvement | –17.3 | 3.7 | –21.0 |
| **30 to 90** | Stage 3 | 2075.4 | 1864.1 | 211.4 |
| | Stage 4 | 2070.0 | 1871.6 | 198.4 |
| | Improvement | –5.4 | 7.6 | –13.0 |
| **25, 30 to 90** | Stage 3 | 2319.9 | 1912.9 | 407.0 |
| | Stage 4 | 2301.7 | 1920.9 | 380.8 |
| | Improvement | –18.2 | 7.9 | –26.2 |
| **20 to 90** | Stage 3 | 2472.8 | 1997.0 | 475.8 |
| | Stage 4 | 2451.3 | 2002.7 | 448.6 |
| | Improvement | –21.5 | 5.7 | –27.2 |

Notes: PVUC at the unconstrained optimum has been normalised to zero with PVTTC = PVAC + PVUC in all rows. Hence, PVUC values in for constrained scenarios are the excess over the PVUC value at the unconstrained optimum. The 'improvement' is the stage 4 minus and stage 3 present value result.

Table 6.9 presents the present values for the 20-year uniform budget constraints in the same format as Table 6.4 for the 10-year constraints. Both tables show PVAC falling up to a point and then rising as the constraints were tightened. This is because, with tight constraints, the additional road agency spending required following the end of the constrained years to repair the neglect of the constrained years is high enough to exceed the savings from the budgets even after discounting. Again, the IBCR for each step is between the average of the annual MBCRs for the start and end constraint points over the constrained period, with the exception of the IBCR comparing the unconstrained and $80 million constrained scenario.

**Table 6.9 Present values and BCRs for 20-year uniform annual budget constraints: Stage 4 optimisation**

*($ millions)*

| | Constraint | PVTTC | PVAC | PVUC | PVB | IBCR | Avg MBCR |
|---|---|---|---|---|---|---|---|
| | nil | 1902 | 1902 | 0 | 1084 | | 1.0 |
| **Δ** | **–** | **20** | **–31** | **51** | **–24** | **1.8** | |
| | 80 | 1922 | 1871 | 51 | 1060 | | 1.1 |
| **Δ** | **–20** | **111** | **–59** | **170** | **–245** | **1.5** | |
| | 60 | 2033 | 1812 | 221 | 815 | | 1.7 |
| **Δ** | **–10** | **239** | **11** | **228** | **–136** | **2.8** | |
| | 50 | 2272 | 1823 | 449 | 679 | | 4.2 |
| **Δ** | **–1.6** | **131** | **34** | **97** | **–22** | **7.1** | |
| | 48.4 | 2403 | 1857 | 547 | 658 | | 7.7 |

Notes: The values of PVTTC were calculated with PVUC set to zero at the unconstrained optimum as in Table 6.8.
PVB is the present value of road agency spending over the 20 constrained years after the last stage of optimisation.
The rows headed Δ show differences between the rows just above and just below.
The IBCRs are comparisons between the rows just above and just below, not the first row with unconstrained spending.
The average MBCR is the simple average of the MBCRs for the 20 budget-constrained years.

## 6.6 Case study results: minimising agency costs with annual budget constraints

The next set of optimisations subject to annual budget constraints was for the cost–effectiveness analysis approach, minimising PVAC subject to maximum roughness constraints.

In the first stage of the optimisation process, testing all options, the maintenance model was modified to minimise PVAC subject to maximum roughness constraints, discarding options that fail to meet the maximum roughness constraint. The remaining three stages — removal of dominated options, imposition of annual budget constraints with penalty factors, and solution refinement with a genetic algorithm — were the same as for PVTTC minimisation subject to annual budget constraints, but minimising PVAC and working only with options that met the maximum roughness constraint.

After stage 2, there were 115 696 non-dominated options for the entire database, an average of 56.9 options per segment, with numbers of non-dominated options for individual segments ranging between 1 and 177.

Table 6.10 presents the results for PVAC minimisation with 20-year annual budget constraints for four scenarios with the unconstrained spending results (that were shown in Figure 5.6) in the first row of spending amounts. The first set of results, for a uniform $68 million annual budget, left no spike in spending after year 20 and so might be considered at or above a sustainable level. The $61.8 million constraint for all 20 years was the lowest uniform budget constraint that could be achieved while keeping pavements within the maximum roughness constraints. It seems that, compared with PVTTC minimisation, the loss of options caused by the maximum roughness constraints reduced the flexibility of the model to shift treatments between years and achieve lower annual budget targets.

The other two scenarios involved tightly-constrained budgets in the early years compensated by higher spending several years later. Spending less in the early years raised the minimum achievable levels of spending in subsequent years. In the case of the $61.8 million uniform constraint, scenario B, spending was pushed into the year 20 and beyond. For scenarios C and D, the very tight constraints in the first few years made it impossible to reduce constraints a few years later to below $95.5 and $120 million respectively. However, spending needs reduced by year 17 or earlier.

**Table 6.10 Model results for 20-year annual budget constraints minimising agency costs with maximum roughness levels**

*($ millions)*

| Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Unconstrained spending | 161.5 | 33.0 | 49.2 | 58.3 | 80.1 | 52.0 | 12.5 | 37.0 | 57.4 | 79.5 | 51.7 | 59.2 | 102.7 | 72.7 | 43.4 | 67.0 | 68.8 | 56.7 | 34.1 | 60.5 | 52.8 | 52.4 |
| **Scenario A** | | | | | | | | | | | | | | | | | | | | | | |
| Constraints | 68 | | | | | | | | | | | | | | | | | | | | nil | |
| Spending 3 | 68.0 | 67.9 | 67.7 | 67.9 | 67.9 | 65.7 | 67.6 | 61.5 | 65.5 | 63.1 | 67.4 | 66.2 | 61.9 | 66.6 | 57.6 | 68.0 | 68.0 | 62.7 | 29.5 | 57.2 | 53.7 | 48.5 |
| Spending 4 | 68.0 | 68.0 | 68.0 | 68.0 | 68.0 | 68.0 | 64.5 | 61.4 | 59.8 | 67.7 | 68.0 | 64.4 | 65.2 | 67.3 | 56.8 | 68.0 | 68.0 | 62.9 | 29.1 | 57.7 | 52.9 | 48.4 |
| Lambdas | 0.7 | 0.5 | 0.4 | 0.3 | 0.2 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| MBCRs | 1.7 | 1.6 | 1.4 | 1.3 | 1.2 | 1.1 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| **Scenario B** | | | | | | | | | | | | | | | | | | | | | | |
| Constraints | 61.8 | | | | | | | | | | | | | | | | | | | | nil | |
| Spending 3 | 61.8 | 61.8 | 61.8 | 61.8 | 61.8 | 61.5 | 61.7 | 61.4 | 61.8 | 57.2 | 59.4 | 60.3 | 47.5 | 56.4 | 59.7 | 51.8 | 46.2 | 48.1 | 61.1 | 56.1 | 231.8 | 117.9 |
| Spending 4 | 61.8 | 61.8 | 61.8 | 61.8 | 61.8 | 61.8 | 61.8 | 61.5 | 61.8 | 61.8 | 61.8 | 61.8 | 61.7 | 61.8 | 61.8 | 61.8 | 61.8 | 61.8 | 60.6 | 61.7 | 172.2 | 104.4 |
| Lambdas | 10.1 | 8.2 | 7.0 | 5.9 | 5.1 | 4.3 | 3.0 | 2.6 | 2.4 | 2.3 | 2.2 | 1.9 | 1.8 | 1.6 | 1.3 | 1.3 | 1.4 | 1.2 | 0.4 | 0.3 | 0.0 | 0.0 |
| MBCRs | 11.5 | 9.9 | 8.9 | 7.9 | 7.2 | 6.4 | 5.0 | 4.6 | 4.5 | 4.4 | 4.4 | 4.0 | 4.0 | 3.8 | 3.4 | 3.5 | 3.7 | 3.4 | 1.8 | 1.6 | 1.0 | 1.0 |
| **Scenario C** | | | | | | | | | | | | | | | | | | | | | | |
| Constraints | 35 | 45 | 55 | 65 | 75 | 85 | 95 | | | | | | | | | | | | | | nil | |
| Spending 3 | 34.7 | 45.0 | 54.1 | 64.9 | 72.1 | 83.7 | 94.5 | 94.8 | 94.1 | 91.5 | 93.2 | 95.0 | 92.7 | 94.6 | 93.4 | 90.4 | 62.7 | 56.8 | 38.9 | 45.9 | 43.5 | 40.0 |
| Spending 4 | 35.0 | 45.0 | 55.0 | 65.0 | 75.0 | 85.0 | 95.0 | 95.0 | 95.0 | 95.0 | 95.0 | 94.9 | 94.6 | 94.8 | 80.6 | 94.6 | 57.0 | 54.2 | 32.9 | 45.6 | 40.9 | 43.1 |
| Lambdas | 18.1 | 8.3 | 6.7 | 4.8 | 3.6 | 2.5 | 1.5 | 1.0 | 0.7 | 0.6 | 0.4 | 0.3 | 0.2 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| MBCRs | 19.9 | 10.0 | 8.5 | 6.6 | 5.4 | 4.1 | 3.0 | 2.4 | 2.0 | 1.9 | 1.7 | 1.5 | 1.3 | 1.2 | 1.1 | 1.1 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| **Scenario D** | | | | | | | | | | | | | | | | | | | | | | |
| Constraints | 35 | 35 | 35 | 60 | 90 | 120 | | | | | 110 | | | | | | | | | | nil | |
| Spending 3 | 35.0 | 34.9 | 33.7 | 60.0 | 90.0 | 120.0 | 118.2 | 117.6 | 120.0 | 118.3 | 101.6 | 109.6 | 108.6 | 84.5 | 74.5 | 95.9 | 54.0 | 50.6 | 29.8 | 24.6 | 33.5 | 37.6 |
| Spending 4 | 35.0 | 35.0 | 35.0 | 60.0 | 90.0 | 120.0 | 120.0 | 120.0 | 119.9 | 120.0 | 109.6 | 110.0 | 109.7 | 96.5 | 59.7 | 96.8 | 47.4 | 43.5 | 25.2 | 19.4 | 25.0 | 38.8 |
| Lambdas | 22.0 | 15.8 | 14.7 | 6.1 | 4.5 | 2.5 | 1.7 | 1.0 | 0.6 | 0.5 | 0.3 | 0.2 | 0.1 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| MBCRs | 23.8 | 18.1 | 17.5 | 8.1 | 6.5 | 4.2 | 3.2 | 2.4 | 1.9 | 1.7 | 1.5 | 1.3 | 1.2 | 1.2 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |

Notes: Spending 3 is results from the stage 3 optimisation using penalty factors. Spending 4 is results after the stage 4 optimisation letting the genetic algorithm select individual options. Although the budget constraints were imposed for years 1 to 20 only, years 21 and 22 were included the table to show how spending was pushed out into the first few unconstrained years.

Section 3.5 on PVAC minimisation subject to road roughness constraints showed that the MBCRs derived from the penalty factors with the formula $(1 + r)^t \lambda_t + 1$ indicate the net financial benefit to the road agency from spending an additional dollar of present value in a budget-constrained year (not counting the increase in PVAC from the extra dollar).

Table 6.11 shows the present values for stages 3 and 4 with the improvements from the final stage. The improvements saved costs for both the road agency and road users, unlike PVTTC minimisation where the final stage increased road agency costs. Nevertheless, the changes in the present values were small.

Comparing the PVAC minimisation results for the $61.8 million uniform annual budget constraints with the PVTTC minimisation results for the $60 million uniform annual budget constraints over 20 years, the former cost society an additional PVTTC of $368 million. This is comprised of $485 million of higher user costs, offset by a $116 million saving in PVAC.

**Table 6.11 Impacts on present values of 20-year annual budget constraints and of stage 4 optimisation for PVAC minimisation**

*($ millions)*

| Constraints | Optimisation stage | PVTTC | PVAC | PVUC |
|---|---|---|---|---|
| **Nil** | | 1623.1 | 1623.1 | 0.0 |
| **$68m uniform** | Stage 3 | 1644.3 | 1640.8 | 3.4 |
| | Stage 4 | 1643.8 | 1640.7 | 3.1 |
| | Improvement | –0.5 | –0.2 | –0.3 |
| **$61.8 uniform** | Stage 3 | 1749.4 | 1702.9 | 46.5 |
| | Stage 4 | 1737.7 | 1695.4 | 42.3 |
| | Improvement | –11.7 | –7.5 | –4.3 |
| **$35 m rising to $95m** | Stage 3 | 1773.8 | 1768.2 | 5.6 |
| | Stage 4 | 1762.3 | 1762.3 | 0.0 |
| | Improvement | –11.5 | –5.9 | –5.6 |
| **$35 m for 3 years rising to $120m** | Stage 3 | 1789.1 | 1829.7 | –40.5 |
| | Stage 4 | 1777.8 | 1825.7 | –47.9 |
| | Improvement | –11.3 | –4.0 | –7.3 |

Notes: PVUC at the unconstrained optimum has been normalised to zero with PVTTC = PVAC + PVUC in all rows. Hence, PVUC values for constrained scenarios are the excess over the PVUC value at the unconstrained optimum. The 'improvement' is the stage 4 minus the stage 3 present value result.

## 6.7 Simple triage method for year-one treatments

An important lesson from the case study is that a substantial proportion of the first-year backlog can be deferred at little cost to road users and the road agency. The cost to society can be kept to a minimum by shifting treatments out of year one in ascending order of the cost of the deferral per dollar of year-one spending saved, until the budget constraint is met. The stage 3 optimisation methodology, using penalty factors, demonstrated in this chapter, does just that.

A simple method to identify year-one treatments that can be deferred at little cost to society, and that can be readily applied by practitioners, is to apply a penalty factor to costs incurred in year one only. The practitioner still needs to start by identifying and costing a large number of options with different treatments at different times for all segments in the network. Then, starting with a small penalty factor for year one only, for each segment $i$, select options to minimise $PVTTC_i + \lambda_1 c_{i1}$, where $\lambda_1$ is the year-one penalty factor and $c_{i1}$ is the cost of a treatment in year one. For segments with no treatment in year one in the optimal solution, there will be no change. Raising the penalty factor will cause some options that, in the optimal solution, have treatments in year one, to be replaced with options with their first treatment in other years, reducing total spending in year one. The penalty factor needs to be increased to the point where the year one budget constraint is met.

Table 6.12 illustrates the method with a simple hypothetical numerical example for a single segment. Four options are shown, labelled *A* to *D*, with PVTTC values increasing in columns further to the right. Options *A* and *B* have treatments in year one. Options *C* and *D* have treatments in years two and three, respectively. They all have treatments in subsequent years not shown in the table, for example a resurface in year 16. Option *A*, having the lowest PVTTC, is the most economic option without budget constraints. The second part of the table shows PVTTC + $\lambda c$ values computed from successive $\lambda$ values rising from 0.1 to 0.4. Option A has the minimum PVTTC + $\lambda c$ value up to a $\lambda$ value of 0.2. Above 0.2, option C has the minimum PVTTC + $\lambda c$. So as the $\lambda$ value is progressively increased, the year-one treatment for the segment will shift out of year one after the $\lambda$ value rises above 2.0.

**Table 6.12 Numerical example illustrating simple triage method**

*($'000)*

| Option | A | B | C | D |
|---|---|---|---|---|
| **PVTTC** | 1000 | 1010 | 1020 | 1030 |
| | | Treatment costs by year | | |
| **Year** | | | | |
| **1** | 100 | 100 | | |
| **2** | | | 110 | |
| **3** | | | | 120 |
| | | Treatment costs in subsequent years not shown | | |
| **$\lambda$ values** | | $PVTTC_i + \lambda\, c_{i1}$ | | |
| **0.1** | 1010 | 1020 | 1020 | 1030 |
| **0.2** | 1020 | 1030 | 1020 | 1030 |
| **0.3** | 1030 | 1040 | 1020 | 1030 |
| **0.4** | 1040 | 1050 | 1020 | 1030 |

Note: Greyed cells are row minimums.

Table 6.13 shows the number of numbers of treatments correctly and incorrectly identified for retaining in year one and for dropping from year one using the simple triage method for each of the PVTTC and PVAC minimisation scenarios in the previous three sections. In the case of $100 million annual budget constraints for 10 years, the simple triage method worked extremely well. Of the 646 treatments in the first year of the optimal solution, with alternative treatment options, only eight treatments were incorrectly excluded and an additional seven treatments were incorrectly added. For the $80 million PVTTC minimisation scenarios, the level of misclassifications was higher but might still be considered acceptable. Thereafter, as constraints were tightened further, the accuracy of the method deteriorated rapidly with the number of mistakes exceeding the number of correctly-identified treatments. As rough guide, Table 6.13 suggests that the simple triage method can only be recommended for penalty factors below roughly 0.15, but more case studies would be needed to confirm this.

MBCRs calculated from the penalty factors under the simple triage method should be disregarded. They are only correct under the assumption of no budget constraints after year one. Imposing a penalty on year-one spending shifts much of the spike in year-one spending into year two, which will be just as unrealistic to implement as the spike in year one.

### Table 6.13 Test results for simple triage method for year-one treatments for a single segment

| Constraint | Penalty factor | Number of treatments in year one [a] | | | | | F-measure [d] |
|---|---|---|---|---|---|---|---|
| | | Optimal solution | Correctly identified (true positive) | Incorrectly added (false positive) | Incorrectly excluded (false negative) | Correct segment with incorrect treatment type [b] | |
| **Column no. [c]** | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| **Unconstrained** | 0 | 1123 | | | | | |
| | | | **PVTTC minimisation: 10-year constraints** | | | | |
| **$100 m uniform** | 0.1243 | 646 | 638 | 7 | 8 | 0 | 99% |
| **$80 m uniform** | 0.1810 | 499 | 442 | 42 | 57 | 0 | 90% |
| **$60 m uniform** | 0.2429 | 339 | 244 | 74 | 94 | 1 | 74% |
| **$50 m uniform** | 0.2690 | 266 | 147 | 99 | 118 | 1 | 58% |
| **$40 m uniform** | 0.3066 | 172 | 59 | 116 | 111 | 2 | 34% |
| **$35.3 m uniform** | 0.3238 | 138 | 40 | 106 | 97 | 1 | 28% |
| | | | **PVTTC minimisation: 20-year constraints** | | | | |
| **$80 m uniform** | 0.1810 | 511 | 448 | 49 | 63 | 0 | 89% |
| **$60 m uniform** | 0.2429 | 358 | 220 | 111 | 137 | 1 | 64% |
| **$50 m uniform** | 0.2690 | 264 | 99 | 159 | 163 | 2 | 38% |
| **$48.4 m uniform** | 0.2769 | 255 | 90 | 153 | 163 | 2 | 36% |
| **$30 m rising** | 0.3565 | 155 | 64 | 54 | 90 | 1 | 47% |
| **$25 m rising** | 0.3854 | 100 | 38 | 63 | 61 | 1 | 38% |
| **$20 m rising** | 0.4682 | 66 | 32 | 35 | 33 | 1 | 48% |
| | | | **PVAC minimisation: 20-year constraints** | | | | |
| **$68m uniform** | 0.1417 | 367 | 332 | 35 | 35 | 0 | 90% |
| **$61.8 uniform** | 0.1665 | 257 | 199 | 105 | 58 | 0 | 71% |
| **$35 m rising to $95 m** | 1.5882 | 99 | 76 | 23 | 23 | 0 | 77% |
| **$35 m for 3 years rising to $120 m** | 1.5882 | 107 | 72 | 27 | 35 | 0 | 70% |

a. Excludes segments where treatment in year one was the sole option. There were 17 such cases for PVTTC minimisation with 10-year constraints, 4 for PVTTC minimisation with 20-year constraints, and 18 for PVAC minimisation with 20-year constraints.

b. There were up to two cases in each row of the table where, for segments with multiple options with treatments in year one, the one-year penalty factor method chose the wrong treatment.

c. Column 2 = column 3 + column 5 + column 6

d. The F-measure or F-score is used to evaluate binary classification algorithms that classify instances into 'positive' or 'negative'. It is commonly used for evaluating information retrieval systems and machine learning models. The F-measure is the harmonic mean of $Precision = \frac{TP}{(TP+FP)} = \frac{col.\ 3}{(col.\ 3+col.\ 4)}$ and $Recall = \frac{TP}{(TP+FN)} = \frac{col.\ 3}{(col.\ 3+col.\ 5)}$, that is $\frac{2}{\frac{1}{Precision}+\frac{1}{Recall}}$.

The simple triage method could be applied to all options from the stage 1 full enumeration without undertaking the stage 2 removal of non-dominated options. Having cut down year-one spending, the method could then be applied to identify treatments to defer in year two and so on. However, the number of errors would increase, and the size of the required spending cut, and hence of the maximum acceptable penalty factor, should be further limited.

## 6.8 Measuring maintenance deficits

Reporting a 'maintenance deficit' draws attention to under-funding of maintenance. This section identifies a number of ways to express maintenance deficits and discusses their advantages and disadvantages. All measures involve comparisons between recent, current or forecast actual spending and future spending needs as forecast by a model. Future spending needs can be estimated at economically-optimal levels or at the level required to meet a set of minimum standards, without or with budget constraints.

### 6.8.1 First year optimal spending

The year-one backlog found by an optimisation without annual budget constraints could be used to indicate of the size of the maintenance deficit. Optimal year-one spending was $186 million in our case study, compared with a $90 million annual average spend over the first 10 years. For minimising PVAC subject to maximum roughness standards, the year-one spending need was $162 million compared with a $62 million average over the first 10 and 20 years. However, a large year-one number can overstate the urgency with which the backlog needs to be addressed. The economic and technical warrant for undertaking the backlog of works in the first year varies from borderline to essential. The $100 million annual budget constraint scenario showed that $86 million of the first-year spending could be deferred to subsequent years for a relatively small cost of a $4 million increase in PVTTC (Table 6.5), with a very low year-one MBCR of 1.2 (Table 6.3). For PVAC minimisation, the $68 million constraint scenario increased PVAC by $18 million (Table 6.11) with a year-one MBCR for the road agency of 1.7 (Table 6.10). The implication is that roughly half of the first-year backlog consists of treatments that could be deferred at a small cost. First-year optimal spending is therefore not recommended as a maintenance deficit measure on the grounds that it could be unduly alarmist.

### 6.8.2 Comparisons with forecast spending

More balanced indicators of spending needs would be based on forecasts for a number of years into the future expressed as an annual sustainable, average annual or the present value of spending needs.

The concept of a 'sustainable' level of spending was defined in Section 6.5 as a constant annual level that does not leave a large backlog at the end of the constrained period or the backlog is not so large that it cannot be caught up by continued spending at the sustainable level in in subsequent years. It was shown that the sustainable levels for our case study network were of the order of $80 million per annum for PVTTC minimisation and $68 million for PVAC minimisation subject to minimum standard constraints. These are somewhat above the average 20-year spending requirements in the absence of budget constraints — $75 million and $62 million respectively. The average spending level therefore is likely to be a more conservative estimate of requirements than the sustainable level.

Tsunokawa and Ul-Islam (2003) defined to the maintenance gap ratio (MGR) as

$$MGR = 1 - \frac{PVAC_{budget}}{PVAC_{opt}}$$

where $PVAC_{opt}$ is the present value of unconstrained economically-optimal maintenance spending by the road agency and $PVAC_{budget}$ is the present value of the maintenance budget over the analysis period. A budget that fully met optimal maintenance needs would have an MGR of zero. The MGR is the proportion of funding needs that will not be met. The more constrained the budget, the higher the MGR, up to a maximum of one where no maintenance is undertaken at all.

Estimation of the MGR requires assumptions to be made about the size of future budgets over the entire optimisation period, which was 40 years for Tsunokawa and Ul-Islam (2003, p. 197) — a very long-range forecast. The period over which the present values are calculated could be shortened to say 10 or 20 years. To avoid the need for assumptions about budgets in the distant future, we suggest defining the MGR as

$$MGR = 1 - \frac{current\ or\ forecast\ annual\ spending}{average\ annual\ sustainable\ spending}$$

In our case study, if $80 million was the sustainable level, as proposed in Section 6.5. If $30 million was available to spend each year, the MGR would be 62.5%.[31] The maintenance gap could also be reported simply as the money amount by which the annual current spending level falls short of the sustainable level.

## 6.8.3 Marginal benefit–cost ratio

If likely annual spending could be specified for a number of years into the future, such as the first 10 or 20 years, annual MBCRs can be forecast. As discussed in Chapter 3, MBCRs show the economic value to society of putting additional funds into maintenance in specific years and can be compared with benefit–cost ratios for capital spending. IBCRs can also be used indicate the value of a specified spending increase in one or more years compared with the alternative. While funding to the level of the economic optimum might be considered an unrealistic goal, there is a strong case for funds to be allocated to achieve an MBCR for maintenance the same as for capital investment, as explained in Chapter 3. The amount of maintenance spending consistent with an average MBCR over the constrained period equal to the MBCR for investment spending could be made the standard of comparison to indicate the extent of a maintenance deficit.

## 6.8.4 Equivalent interest rate for deferred maintenance

Naudé et al. (2008) and (2012) used a case study to demonstrate that, in a situation of tight annual budget constraints, savings in the present values of road agency and road user costs can be made by bringing forward maintenance interventions. Under the ‘Accelerated Road Rehabilitation Program’, the Queensland state government’s treasury department provided loans to the road agency, which the road agency repaid with interest out of future budget allocations. There was a net saving to the road agency as well as to road users. The case study illustrates that, in the long term, governments may be better off borrowing funds to avoid deferring maintenance.

Deferring maintenance can be seen as a form of borrowing. Funds are saved in the short-term at the expense of higher outlays in the future. A way to communicate this to decision makers and enable comparison with the alternative of borrowing to fund maintenance would be to estimate an ‘equivalent interest rate for deferred maintenance’ (EIRDM).

It is assumed that funds are constrained to the point where only minimum acceptable pavement conditions can be provided. The required modelling would be the same as for our case study for minimising PVAC subject to maximum roughness constraints. To use higher standards to estimate the EIRDM might be misleading because, if roads were allowed to deteriorate below the specified higher minimum standards and never restored to those standards, the equivalent loan would never be repaid in the sense of the road agency having to spend more later to compensate for spending less in the short term. The minimum standards could be set using expert judgement with the aim of ensuring each road is just able to fulfil its economic and social purposes. Since the objective is to minimise costs to the government rather than to maximise economic efficiency, the interest rate at which the government can borrow should be used as the discount rate in the optimisation model rather than the social discount rate.

Table 6.14 uses the forecast spending levels in Table 6.10 to illustrate calculation of EIRDMs. Scenarios A and B are taken as two alternative sustainable constant levels of future spending and scenarios C and D as spending profiles that save funds in the early years but require much higher levels of spending in later years to avoid any segments falling below the minimum roughness levels.

The four columns show differences for the four possible comparisons between the constant and variable spending scenarios. They were derived simply by subtracting the spending after stage 4 optimisation in the rows in Table 6.10 for scenarios C and D, from the spending after the stage 4 optimisation for scenarios A and B. For example, in year one, by choosing scenario C over scenario A, the road agency spends $35 million compared with $68 million, a saving of $33 million. In year two, the saving is $68 million – $45 million =

----------

[31] Provided the annual spending amounts are constant over time, the annual amounts can be entered into the MGR formula because the present values will be proportional to the annual amounts.

$23 million. By year five, spending under scenario C has risen above spending under scenario A so the saving is $68 million – $75 million = –$7 million. In effect, from year five on, the road agency is repaying its borrowings. The negative net cash flows continue in all four cases up to year 17 when the maintenance backlog is fully caught up.

**Table 6.14 Differences between spending under scenarios in Table 6.10 and internal rates of return**

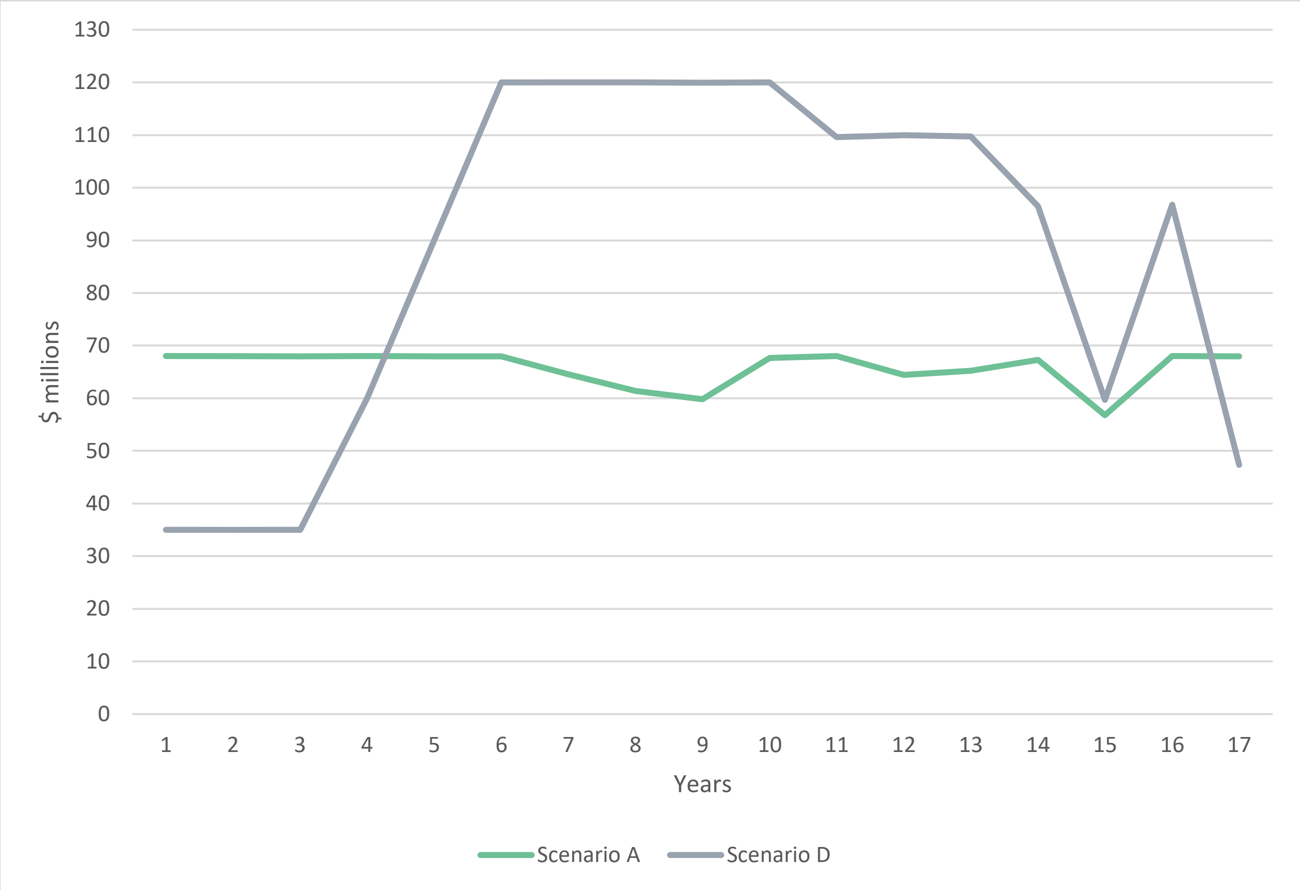
| | A – C | A – D | B – C | B – D |
|---|---|---|---|---|
| **Internal rate of return over 40 years** | 19.2% | 24.4% | 27.1% | 31.2% |
| **Internal rate of return over first 17 years** | 19.4% | 24.8% | 28.1% | 31.7% |
| **Year** | **Cash flow ($ millions)** | | | |
| **1** | 33 | 33 | 27 | 27 |
| **2** | 23 | 33 | 17 | 27 |
| **3** | 13 | 33 | 7 | 27 |
| **4** | 3 | 8 | –3 | 2 |
| **5** | –7 | –22 | –13 | –28 |
| **6** | –17 | –52 | –23 | –58 |
| **7** | –30 | –55 | –33 | –58 |
| **8** | –34 | –59 | –34 | –59 |
| **9** | –35 | –60 | –33 | –58 |
| **10** | –27 | –52 | –33 | –58 |
| **11** | –27 | –42 | –33 | –48 |
| **12** | –31 | –46 | –33 | –48 |
| **13** | –29 | –44 | –33 | –48 |
| **14** | –28 | –29 | –33 | –35 |
| **15** | –24 | –3 | –19 | 2 |
| **16** | –27 | –29 | –33 | –35 |
| **17** | 11 | 21 | 5 | 14 |
| **18** | 9 | 19 | 8 | 18 |
| **19** | –4 | 4 | 28 | 35 |
| **20** | 12 | 38 | 16 | 42 |
| **21** | 12 | 28 | 131 | 147 |
| **22** | 5 | 10 | 61 | 66 |

The table shows the internal rates of return (IRRs) for the four cash flows for all 40 years of the analysis period, including depreciation at the end of year 40, and for the first 17 years just after the period of negative cash flows ends. Whether it is calculated over the first 17 years or the whole analysis period makes little difference to the IRR.

The IRR is the discount rate that makes the present value of the cash flows zero. In this case, it shows the interest rate the road agency would pay if it maintained spending at a constant sustainable level and borrowed to achieve the same cash flow as under a variable spending scenario. The lender would earn interest of the order of 20% to 30%, which is extremely high. If the road agency were to maintain spending at a constant level and borrow at the interest rate available to the government, it could achieve the same cash flow in the early years and have much less to pay back later. This type of analysis could be used by a road agency to justify government borrowing to fund maintenance in the short term.

Figure 6.4 plots the spending amounts under scenarios A and D from Table 6.10. The gap between the two curves for the first four years is the amount gained by the road agency by adopting scenario D compared with scenario A (A – D). The gap between the two curves from years five to 17 is the amount forgone. To make the present value of these two areas, one positive and the other negative, equal to zero, the discount rate has to set at 24.8%. At any lower discount rate, the loss from years five to 17 exceeds the gain from years one to four producing a negative net present value. If the road agency adopted scenario A and borrowed during years one to four to make spending in those years the same as for scenario D, as long as interest rate was below 24.8%, it would experience lower cash outflows from years four to 17 while repaying the loan compared with scenario D. In other words, so long as the road agency can borrow at an interest rate below 24.8%, it will be better off following scenario A and borrowing to fund early maintenance within budget constraints.

**Figure 6.4 Spending under scenarios A and D up to year 17**




## 6.9 Conclusion

The case study of optimising maintenance subject to annual budget constraints has shown that the four-stage methodology developed can cope with a large database of road segments. Indeed, as long as computational limits are not exceeded, a larger database is better because the discrete optimisation problem bears a closer resemblance to a continuous problem, and so is more amenable to application of the penalty method.

The difficulty of the task increases with the number of years for which budget constraints are imposed because fewer options can be eliminated as dominated in stage 2 of the optimisation and there are more penalty factors to set in stage 3. However, a short period of budget constraints can be unrealistic where it results in a large backlog of expenditure in the year immediately following the constrained period.

Stage 4 of the optimisation makes only a small improvement in terms of the objective function while more fully utilising available budgets. The contribution of stage 4 is greater for tighter budget constraints because the penalty method has less flexibility in stage 3 to shift treatments between years to achieve the best outcome. In practical applications, stage 4 can reasonably be omitted, except in cases of very tight budget constraints or a relatively small number of segments in the database leading to significant lumpiness in options.

From the penalty factors for each year, annual MBCRs can be obtained using the formula given in Chapter 3. With uniform and rising annual budget constraints, annual MBCRs tend to be highest in the first year, when the demand for funds is greatest, then progressively decline.

As annual constraints were tightened, PVAC fell initially, and then rose as the additional costs in later years to recover from the underspending in early years predominated. Progressive tightening of annual budget constraints at first required only small penalty factors, and hence annual MBCR estimates, but then rose rapidly.

Optimisation subject to loose budget constraints, for example, set at a level around or somewhat above the 10-year or 20-year average annual optimal spending in the absence of budget constraints, shows that a large portion of the year one backlog can be deferred at only a small economic cost. This is because the stage 3 optimisation process, using penalty factors, selects treatments to defer on the basis of the lowest cost in terms of the objective function.

This leads to the idea of a simple triaging method to prioritise year-one treatments for deferral in order to meet a budget constraint. Application of a penalty factor for the first year alone can be used to identify treatments to defer that fairly accurately accord with the recommendations arrived at by application of the penalty method over a number years, provided that the budget constraint is not too tight. The case study suggested the method works satisfactorily for constraints up to roughly half the first-year backlog or a penalty factor not exceeding 0.15.

A number of ways to measure maintenance deficits were considered. The year-one backlog can be misleading as a maintenance deficient measure because a major portion of the backlog can be deferred at little cost. Better guides are comparisons with a sustainable level of spending or optimal spending averaged over 10 or 20 years, compared with current or forecast spending. Annual MBCRs are useful. The ‘equivalent interest rate for deferred maintenance’ (EIRDM) can be used to convey to decision makers that deferring road maintenance to save funds in the short term is akin to borrowing money that has to be repaid later, but it can be very expensive. It is better to keep maintenance at a sustainable level by borrowing at the interest rate available to the government and repay it later.

# 7. Review in light of study objectives

This final chapter of the report summarises the main contributions of the report and recaps it with regard to the five objectives set out in the introduction.

## 7.1 Report's contributions

The voluminous literature on road maintenance optimisation comes almost entirely out of the civil engineering discipline. While drawing heavily on that literature, the present report addresses road maintenance from the perspective of the economics discipline. The report focuses only on the cost minimisation objective rather than the non-monetary objectives found in the engineering literature such as weighted area under a performance curve. It applies Lagrange's method for optimisation subject to constraints, which is very familiar to economists. Only one article in the literature has used this approach to optimise subject to annual budget constraints.

A road maintenance optimisation exercise with a long analysis period and a large number of road segments to consider together faces the 'curse of dimensionality', due to the astronomically large number of possible options with different treatment types in different years. The principle ways in which the case study in this report has managed the curse of dimensionality has been first, excluding all options with implausibly short intervals between treatment times, and second through a multi-stage optimisation process. The pavement optimisation literature to date discusses only two-stage approaches, the first stage being to identify the best one or more treatment options for each segment individually, and the second stage, to select options to fit within annual budget constraints. In the literature to date, the second stage of the optimisation has been undertaken using prioritisation methods, which are not guaranteed to find the best possible solution. The present report, therefore, has introduced some significant innovations in road maintenance optimisation enabling large databases of segments to be optimised subject to annual budget constraints.

The economic value of capital projects can be assessed using cost–benefit analysis (CBA), which produces the widely-recognised measures of net present value (NPV) and benefit–cost ratio (BCR). Road maintenance does not fit well into the CBA framework that compares a do-minimum base case with one or a small number of project options. For each road segment, there are alternative treatment types that can be applied with different intensities and that can be implemented in different years, giving rise to a huge number of potential treatment options. There is arbitrariness in choosing the do-minimum base case. The maintenance optimisation literature does not provide a way to indicate the economic value of spending on maintenance comparable with the BCR of a construction project. The present report has proposed the marginal benefit–cost ratio (MBCR) for maintenance. The MBCR for a given year indicates the economic value of additional maintenance funding in that year and is comparable with BCRs for capital projects. The report sets out the economic interpretation of the values of Lagrange multipliers or penalty factors at the constrained optimum for a number of segments optimised together and shows how to convert them to MBCRs.

The report discusses ways in which the extent of underfunding of maintenance can be assessed and brought to the attention of decision makers. The first-year backlog of spending estimated by a maintenance model is not recommended because it is likely that a significant proportion of that expenditure can be deferred at only a small economic cost. Forecast annual average or sustainable spending needs over the next 10 or 20 years are preferable and can be compared with current and expected future annual spending levels. A 'sustainable' level of spending was defined as one where there is no jump in optimal spending just following the constrained period, or where the jump is not so large that it cannot be caught up by continued spending at the sustainable level in subsequent years.

The equivalent interest rate for deferred maintenance (EIRDM) concept highlights the cost of deferring essential maintenance. It requires specification of the lowest acceptable conditions for the network. Spending at unsustainably low levels over the first few years of the analysis period necessities greater spending in later years to keep the network at the lowest acceptable condition. Saving funds in the short term in exchange for higher outgoings in the long term is akin to borrowing. However, borrowing by skimping on maintenance can be expensive because the additional costs incurred later far outweigh the short-term savings even with discounting. Taking the difference in cash flow profiles between a spending scenario with a short-term saving followed by higher catch-up spending later, and a scenario with constant annual spending at a sustainable level, and estimating the internal rate of return for the difference gives the equivalent interest rate for this from of borrowing. The case study results presented in Chapter 6 suggest it is likely to be well above the rate at which the government can borrow funds. In such cases, deferring maintenance because of short-term budget constraints is short-sighted and is a false economy.

## 7.2 Review of maintenance economics

*Objective: to review the economic principles of road maintenance including the timing, form and quantity of maintenance*

The economics of road maintenance centres on identifying treatment types, times and intensities that minimise the present value of costs to society, subject to technical and budget constraints. Cost minimisation achieves economic welfare maximisation so long as it can reasonably be assumed that changes in road user costs over the relevant range will not lead to changes in traffic levels or vehicle mixes. Since travel decisions are made on the basis of whole-of-trip costs and a whole trip uses numerous road segments, this will usually be a satisfactory assumption.

Economic optimisation of road maintenance involves trading off savings in road agency costs from less intense maintenance treatments undertaken less frequently, against resultant higher costs to road users. The optimisation problem is to minimise the sum of the present value of road user costs (PVUC) and the present value of road agency costs (PVAC), referred to as the present value of total transport costs (PVTTC). PVAC increases as more maintenance is undertaken and PVUC decreases, giving rise to a U-shaped PVTTC curve, obtained by summing the PVAC and PVUC curves. The minimum point of the U is the optimum.

The report also investigated a cost-effectiveness approach preferred by some road agencies whereby PVAC is minimised subject to road condition not falling below exogenously set levels.

The report has focused exclusively on periodic maintenance of sealed roads with flexible pavements. Routine maintenance is left to be treated as an annual spending requirement per square metre or lane-kilometre of road. For periodic maintenance treatments, optimisation models are used to select treatment types and timings either for individual road segments considered alone or for networks of segments considered together. A wide range of model simplifications and optimisation techniques have been applied to address 'the curse of dimensionality' in maintenance optimisation.

Road maintenance optimisation models typically find a large amount of spending to be warranted in the first year of the analysis period to catch up on the backlog, followed by smaller amounts in succeeding years. The amounts in succeeding years can then fluctuate widely. Optimisation subject to annual budget constraints is usually needed reduce and smooth out annual spending to fit within funding and physical input constraints.

Key points about the economics of road maintenance highlighted in the report are as follows.

- Higher road user costs (both cost per vehicle and the number of vehicles) justify higher standards of maintenance, hence it is economically warranted to keep higher trafficked roads in better condition.
- Higher costs of maintenance treatments justify lower maintenance standards.
- Higher discount rates justify deferring maintenance leading to an optimum with undiscounted higher road user costs and lower undiscounted road agency costs.
- The length of the analysis period can affect treatment selection by a model. The effect can be reduced by choosing an analysis period that goes well beyond the period of interest (the 'focus period') and by including, with agency costs, a condition-related residual value or depreciation cost at the end of the analysis period. A condition-related depreciation cost or residual value to a certain extent simulates the omitted PVTTC from the time after the end of the analysis period to infinity. Sensitivity tests involving shorter analysis periods showed that depreciation is not a very good substitute for a longer analysis period. There ought to be some years between the focus period and the end of the analysis period to dampen the effect of the highly approximate nature of depreciation amounts.

- If treatment intensity is made a continuous variable in modelling (for example, overlay thickness), a corner solution is likely with the model choosing either the costliest, most effective treatment intensity or the least costly, least effective treatment intensity permitted. The latter would be applied as frequently as allowed. If a large number of discrete treatments are specified with different trade-offs between implementation cost and effectiveness in improving pavement condition, it is likely that some or many treatment types will never be selected because they are 'dominated' by others with better cost–effectiveness trade-offs. It is therefore preferable to have a small number of treatment type alternatives in the model, not only from the computational point of view.
- Light budget constraints can be imposed with little economic cost provided treatment times and types are optimised within those constraints. A substantial part of the year-one backlog may be deferred for a short time at a small economic efficiency cost provided treatments are prioritised for deferral in ascending order of impact on PVTTC per dollar of treatment cost. However, the economic costs of imposing budget constraints rise steeply, at a more than an exponential rate, as the constraints are progressively tightened.
- Constraining spending in some years pushes maintenance expenses into other years. A period of tight annual budget constraints, say for 10 years, will lead to a large spike in economically warranted spending in the year immediately after the constrained period, for example, in year 11 after 10 years of constraints.
- The proverb 'a stitch in time saves nine' applies. Constrained spending that causes maintenance to be deferred, can increase road agency costs in both undiscounted and present value terms by more than the amount saved in the constrained years.
- Optimising pavement strength together with maintenance, only discussed at the theoretical level in the report, involves the same principles as for road user and road agency costs in maintenance. Stronger pavements save future maintenance costs and conversely, giving rise to a U-shaped total cost curve. If maintenance funds are scarcer than capital funds, while the 'first-best' solution is to equalise the scarcity to obtain similar MBCRs for capital and maintenance spending, a 'second-best' solution is to use some of the less scarce capital funds to build stronger pavements.
- Optimal incentives in maintenance contracts, also addressed only at a theoretical level in the report, require the contractor's remuneration to vary negatively with road user costs, which internalises the additional costs to users of driving on rougher pavements and the savings to users of driving on good quality pavements.

## 7.3 Strategic-level needs assessment

*Objective: to identify an effective approach for assessing current and future spending gaps in road maintenance at a strategic level*

The report developed a methodology that was applied to a large database of 2034 road segments.

Data on the network to be assessed was obtained at the level of short segments of road, assumed to be homogeneous. Required data for individual segments included

- Traffic — AADTs and growth rates by vehicle type (cars and several categories of heavy vehicles)
- Climate / environment — mean monthly precipitation, Thornthwaite moisture index
- Road characteristics — segment length, carriageway width, paved area including shoulders, pavement type, divided or undivided
- Road condition — surface age, percent cracked, pavement age, design and current pavement strength, rut depth and roughness.

Design pavement strengths were unavailable. It was assumed that each road segment was designed to Austroads standards given the projected number of equivalent standard axle loads from heavy vehicles over the 30 years following the last rehabilitation or reconstruction given by the pavement age.

Modelling was undertaken over a 40-year analysis period with a condition-related depreciation amount charged at the end of the analysis period. Three treatment types were specified — resurface, resurface with shape correction, and rehabilitation. To reduce the number of possible treatment options to a manageable level, a minimum time interval of eight years between treatments was assumed. This was well below the time taken for new bitumen to begin to oxidise and crack, assumed to be 12 or 16 years depending on the type of pavement.

The pavement deterioration model was a simplified version of the model within the World Road Association (PIARC) HDM-4 model, and the relationship between road user costs and roughness was taken from ATAP (2016). Good calibration of a maintenance model is essential. Values for HDM-4 calibration coefficients and assumptions for the three treatment types (cost per square metre and effects on pavement condition) were supplied by ARRB.

Optimisation subject to annual budget constraints for the report's case study was done via a four-stage process.

1. Full enumeration of all options — up to 581,485 options for each segment given the three treatment types, a 40-year analysis period and an eight-year minimum time interval between treatments.
2. For each segment elimination of 'dominated' options. Dominated options can never appear in an optimal solution because they can be replaced with another option that yields a lower PVTTC with no additional spending in budget constrained years.
3. Imposition of annual budget constraints though minimising the sum of PVTTC and treatment costs multiplied by a penalty factor for each budget constrained year. Finalising the solution required a genetic algorithm to set the penalty factors to their lowest possible values while keeping spending within the constraints.
4. Refinement of the solution using a genetic algorithm.

## 7.4 Case study

*Objective: to undertake a case study to develop and test the identified methodology*

The case study was undertaken with data on 1977 kilometres of non-urban highways supplied by an Australian state government road agency. ARRB was engaged to curate the data into a suitable form for maintenance modelling and to undertake optimisation modelling to minimise PVTTC using the HDM-4 model. They also provided values for model calibration coefficients and advice on assumptions about treatment types, effectiveness and costs.

Compared with case studies in the literature on maintenance optimisation surveyed in Chapter 4, the methodology applied in this report has been able to optimise over an exceptionally large number of segments (2034) and a long analysis period (40 years). The solutions with budget constraints are not guaranteed to be the perfect optimum, but very close.

The model outputted a recommended list of treatment types and times for each segment, with treatment costs and present values of road user and road agency costs. Annual values for road condition measures and user costs could be extracted for all years in the analysis period.

Results were obtained for

- Minimising PVTTC with
  - no budget constraints
  - present value budget constraints
  - uniform annual budget constraints over 10 years
  - uniform and rising annual budget constraints over 20 years
- Minimising PVAC subject to maximum roughness constraints with
  - no budget constraints
  - uniform and rising annual budget constraints over 20 years.

Sensitivity tests were undertaken for changing the discount rate and pavement strength assumptions, and for a variety of changes to model details such as the analysis period.

## 7.5 Directions for national assessment

*Objective: to suggest directions for a comprehensive assessment of maintenance requirements for the national road network*

The approach could be followed for the non-urban parts of the national road network defined as the network of nationally-important road links determined under the *National Land Transport Act 2014*. Data on road condition needs to be assembled and linked with data on traffic, road characteristics and climate. Model calibration is required for different zones and to reflect the maintenance treatment types and costs for different jurisdictions and regions within jurisdictions. Most of the traffic and road condition data exist for national network roads as well as for state arterial roads. Application to local roads would be restricted by lack of data. Local roads comprise some 39% of the length of paved public roads in Australia (BITRE 2021, pp. 114 and 116)[32].

The model would not be applicable to urban roads without significant modifications. Quite different approaches would be required for concrete pavements and bridges.

The model considered only periodic maintenance. Routine maintenance needs to be added on, estimated as a dollar amount per lane kilometre or square metre of pavement. The unit cost will vary with the location.

Modelling requires

- A pavement deterioration relationship or sub-model that forecasts future pavement condition depending on initial pavement condition, time, axle loads and the climate.
- Specification of a small number of treatment types ranging from resurface to rehabilitation, each having a cost per square metre of pavement to implement and effectiveness in restoring pavement condition. The characteristics and costs of treatments will vary between and possibly within jurisdictions. The cost of a given treatment may be greater for pavements in poorer condition.
- A user cost relationship that estimates costs of time, vehicle operation, crashes, and externalities as a function of roughness. The relationship between user costs and road condition needs more empirical research.
- Specification of technical constraints to prevent the model extrapolating outside the realistic range.

The pavement deterioration component of the BITRE model, though a simplified version of the deterioration model in HDM-4, was still quite complex. A very basic deterioration model consisting of just a few equations could be developed and might be satisfactory provided it was well calibrated. An essential feature of a simple model is that roughness increases more rapidly after cracking occurs, which ensures there is an economic justification for resurfacing treatments in the optimisation model. Having a very basic deterioration model would greatly speed up processing times.[33]

Software used in the case study was Mathematica, Excel and Evolver. To process a large number of segments in a reasonable time, the model could be coded in a fast programming language such as Fortran or C++ and set up to process multiple segments in parallel as does Mathematica. To recode the BITRE model from the functional programming language of Mathematica to an imperative programming language would require some changes of approach.

----------

[32] This estimate is based on 2015 statistics, the latest published showing paved and unpaved roads. In the absence of a split for state and local roads into paved and unpaved roads, it was assumed that all unpaved roads were local roads.

[33] Regression analysis of model outputs from the case study in Chapter 5 suggests that a simple annual incremental deterioration model with only cracking and roughness would be $\Delta R = (0.0014\ ACA + 0.029)\ R$, where *ACA* is the percentage of surface cracked and *R* is roughness at the start of the year. The flat and convex parts of the cracking curve could be approximated by $ACA = If[Surface\ age < ICA, 0, 0.5 + k(Surface\ age - ICA)^2]$ where 0.5 is the percentage of cracking at the end of the year just before crack initiation, *ICA* is the time in years to crack initiation, and *k* is a constant. The case study model suggests the constant, *k*, is of the order of 0.3 for sprayed treatment pavements and 0.75 of asphaltic concrete pavements. The upper, concave part of the S-curve for cracking is unnecessary provided cracking is not permitted to go much above 50%. Analysts would need to recalibrate these relationships from local data or from outputs from a sophisticated calibrated model.

Setting the penalty factors at the minimum values that will cause the annual budget constraints to be satisfied is challenging because the problem is non-smooth and is subject to the 'waterbed' effect. A combination of Excel macros and the Evolver genetic algorithm was used to set the penalty factors for the case study, but it should be possible to develop an algorithm within the programming platform used for the model.

As the stage 4 optimisation involved an extremely large number of variables, one for each segment, it was necessary to use Evolver Industrial Edition, which can handle an unlimited number of adjustable variables. Stage 4 could be omitted because the improvement in the solution is likely to be small and may not be warranted given the highly approximate nature of the exercise. In this report, the stage 4 optimisation served as a confirmation that the stage 3 optimisation worked well. Stage 4 may be worthwhile where there are very tight budget constraints or a relatively small number of segments in the database leading to significant lumpiness in the objective function.

If identification of spending needs is only required in broad terms and budget constraints are light to moderate, the stage 1 optimisation alone may suffice. The year-one backlog could then be reduced to some extent without proceeding to further steps by using the simple triage method presented in Section 6.7 that applies the penalty method to the first year only.

## 7.6 Maintenance versus capital expenditure

*Objective: to contribute to understanding the relative merits of expenditure on maintenance of existing infrastructure and investment in new infrastructure*

Investments in new infrastructure are assessed using CBA. The principal results are the NPV and the BCR. Whether the NPV is positive or negative indicates whether the investment is an economically efficient use of resources. The NPV is also used for selecting between mutually exclusive options for the same project, which is consistent with selecting the maintenance option with lowest PVTTC value for each individual road segment. The BCR indicates the ratio of dollars of benefit per dollar of cost (or resources) expended in present value terms. It is used to choose between projects where there is a budget constraint on capital spending. To obtain the maximum sum of benefits for a given capital spending budget, projects would be selected in descending order of BCR until the budget is used up. If capital spending was not too lumpy, the BCR for the last project accepted, or the 'cut-off BCR', could be considered to be the 'marginal BCR' (MBCR) for capital spending. The MBCR is the number of dollars of benefit society obtains from increasing the budget by one dollar. Hence, if the budget was restrictive enough so that no projects could be accepted with a BCR of 2 or less, then investing an extra $1 present value would yield an additional $2 present value of benefits.

The present report has developed the concept of an MBCR for maintenance spending defined as the saving in the present value of user costs (expressed as a positive number) that results from a one dollar increase in PVAC. It can be expressed in terms of PVTTC.

$$MBCR = -\frac{dPVUC}{dPVAC} = -\frac{(dPVUC + dPVAC)}{dPVAC} + \frac{dPVAC}{dPVAC} = -\frac{dPVTTC}{dPVAC} + 1$$

Minimisation of PVTTC subject to a budget constraint expressed as a present value ($PVAC \leq B$) is analytically not much more difficult than minimisation without a budget constraint. One has only to minimise the sum of PVUC and a weight times PVAC for each segment considered in isolation, and adjust the weight as necessary until the budget constraint for spending on all segments combined is just met. At the weighted-optimum solution, the value of the weight is the MBCR.

For optimising subject to annual budget constraints, implementation of the penalty method, the third stage in the optimisation process, provides estimates of MBCRs for increasing funds in particular years. The methodology involves selecting the option for each segment that minimises PVTTC plus the sum of maintenance costs with costs in each year with a budget constraint multiplied by a penalty factor, that is, minimise $PVTTC_i + \sum_{t=1}^{m} \lambda_t c_{it}$ for each segment *i* where *m* is the number years with budget constraints, $\lambda_t$ is the penalty factor for year *t* and $c_{it}$ is the maintenance cost in year *t* for segment *i*. The penalty factors are adjusted upwards above zero to disadvantage options with treatments in years when the demand for funds exceeds the supply. The penalty factor for year *t*, $\lambda_t$, in the optimal solution with all constraints met, can be interpreted as the saving in PVTTC from relaxing the budget constraint in year *t* by one dollar. The saving in

PVTTC from spending an additional dollar of PVAC in year $t$ is $(1+r)^t\lambda_t$. Hence, annual MBCRs can be derived from penalty factors using the relationship

$$MBCR_t = -\frac{dPVTTC}{dPVAC} + 1 = (1+r)^t\lambda_t + 1$$

For large changes in spending in one or more years, an incremental BCR (IBCR) can be estimated from

$$IBCR = -\frac{\Delta PVTTC}{\Delta PVB} + 1$$

where *ΔPVB* is the change in the present value of spending in budget constrained years.

If the MBCR for maintenance in a given year was above the cut-off BCR or MBCR for capital spending, then, additional net benefits could be obtained by shifting funds from the capital to the maintenance budget, and conversely. With the optimal split of funds, the MBCRs would be the same for capital and maintenance spending.

## 7.7 Conclusion

The present report should increase awareness of the importance of road maintenance and understanding of the economics of road maintenance. Future efforts to quantify the economic value of road maintenance will benefit from more up-to-date and comprehensive acquisition and management of data necessary for road maintenance modelling. Further model development and research into calibration values and the relationship between user cost and road condition will also contribute to the capacity for greater application of maintenance modelling. This should enable estimation of marginal and incremental BCRs comparable with BCRs of construction projects, which will support funding decisions. Data and modelling will also enable better estimation of the size of maintenance deficits and the costs of not addressing them.

# Appendix A – Mathematics of present value minimisation subject to budget constraints

This appendix supplements the material in Chapter 3 by presenting the detailed mathematics for minimising PVTTC subject present value and annual budget constraints, and PVAC subject to maximum roughness and annual budget constraints. In each case the envelope theorem is proved, that is, at the optimum, the Lagrange multipliers equal, not only the partial derivatives of the objective function with respect to the constrained variables, but also the total derivatives.

Annual maintenance spending in year $t$, $c_t$, is assumed to be continuous and PVUC and PVAC, and hence PVTTC, are assumed to be continuous functions of maintenance spending in each year, $t$, from one to infinity. This assumption is not realistic for a single road segment but it holds approximately for a large number of small segments considered together.

## A.1 Minimising PVTTC subject to a present value budget constraint

The optimisation problem is Minimise $PVTTC(c_1, c_2, \dots, c_t, \dots)$ subject to $PVAC \leq B$, where $c_t$ is maintenance spending in year $t$ and $B$ is the present value budget constraint.

Since the constraints are inequalities, the Kuhn-Tucker conditions (also known as the Karush-Kuhn-Tucker conditions) for non-linear programming apply (Beavis and Dobbs 1990, p. 54).

The Lagrangian is

$$L = PVTTC(c_1, c_2, \dots, c_t, \dots) + \lambda\left[\sum_{t=1}^{\infty}\frac{c_t}{(1+r)^t} - B\right]$$

where $PVAC = \sum_{t=1}^{\infty}\frac{c_t}{(1+r)^t}$

The first-order conditions for a constrained maximum are

$\frac{\partial L}{\partial c_t} = \frac{\partial PVTTC}{\partial c_t} + \frac{\lambda}{(1+r)^t} = 0$ for all $t = 1$ to $\infty$ (A.1.1)

$\frac{\partial L}{\partial \lambda} = \sum_{t=1}^{\infty}\frac{c_t}{(1+r)^t} - B \leq 0$

$\lambda\left[\sum_{t=1}^{\infty}\frac{c_t}{(1+r)^t} - B\right] = 0$ (A.1.2)

$\lambda \geq 0$

The equation A.1.2 condition allows for situations where the budget is larger than the unconstrained optimum, in which case $\lambda = 0$ and $\frac{\partial PVTTC}{\partial c_t} = 0$. When the constraint is binding, $\lambda$ will have a positive value. In the two-period case, equation A.1.1 for years one and two, gives

$\frac{\frac{\partial PVTTC}{\partial c_1}}{\frac{\partial PVTTC}{\partial c_2}} = 1 + r$ as shown in Figure 3.5.

The change in PVTTC from a series of small changes in spending in each year, such as would occur for a small change in the constraint on PVAC, is

$$dPVTTC = \sum_{t=1}^{\infty}\frac{\partial PVTTC}{\partial c_t}dc_t$$

Substituting in the relationship in equation A.1.1, $\frac{\partial PVTTC}{\partial c_t} = -\frac{\lambda}{(1+r)^t}$

$$dPVTTC = -\lambda \sum_{t=1}^{\infty} \frac{dc_t}{(1+r)^t}$$

A small change in the constraint leads to a series of small changes in $dc_t$ over time as the additional budget is distributed across years to ensure the greatest possible net saving in PVTTC, fulfilling the equation A.1.1 and A1.2 conditions. The present value of these small changes equals the change in PVAC, that is,

$$\sum_{t=1}^{\infty} \frac{dc_t}{(1+r)^t} = dPVAC$$

Hence

$$dPVTTC = -\lambda\, dPVAC$$

$$\lambda = -\frac{dPVTTC}{dPVAC}$$

# A.2 Minimising PVTTC subject to annual budget constraints

Annual budget constraints are imposed for years 1 to *m*. Thereafter, spending is unconstrained for years *m + 1* to infinity.

The problem is Minimise $PVTTC(c_1, c_2, \dots, c_m, c_{m+1}, \dots)$ subject to $c_t \leq B_t$ for all *t* = *1* to *m*

The Lagrangian is

$$L = PVTTC(c_1, c_2, \dots, c_m, c_{m+1}, \dots) + \sum_{t=1}^{m} \lambda_t (c_t - B_t)$$

The Kuhn-Tucker conditions for a maximum subject to inequality constraints are

$$\frac{\partial L}{\partial c_t} = \frac{\partial PVTTC}{\partial c_t} + \lambda_t = 0 \text{ for all } t = 1 \text{ to } m \qquad \text{(A.2.1)}$$

$$\frac{\partial L}{\partial c_t} = \frac{\partial PVTTC}{\partial c_t} = 0 \text{ for all } t = m + 1 \text{ to } \infty \qquad \text{(A.2.2)}$$

$$\frac{\partial L}{\partial \lambda_t} = c_t - B_t \leq 0 \text{ for all } t = 1 \text{ to } m$$

$$\lambda_t (c_t - B_t) = 0 \text{ for all } t = 1 \text{ to } m \qquad \text{(A.2.3)}$$

$$\lambda_t \geq 0 \text{ for all } t = 1 \text{ to } m$$

The condition in equation A.2.3 allows for constraints to be non-binding for some or all constrained years. If, for a given year, the optimal solution does not exhaust all the available budget, then $\lambda_t = 0$ and $c_t - B_t < 0$. Otherwise, where the constraint is binding, $\lambda_t > 0$ and $c_t - B_t = 0$.

From equation A.2.1, $\lambda_t = -\frac{\partial PVTTC}{\partial c_t}$, which implies that the value of lambda for a given year is the saving in PVTTC from a one dollar increase in the budget for the year with spending in all other years held constant. However, from the envelope theorem, $\lambda_t = -\frac{dPVTTC}{dc_t}$ at the optimum, that is, the value of lambda for a given year is the saving in PVTTC from a one dollar increase the budget for the year, after making all consequential optimal adjustments to spending in all other years (Cornes 1992).

To show why this is the case, say the budget constraint for a single year, *i*, is relaxed by one dollar. This leads to changes in spending for other years, so the total change in PVTTC is:

$$\frac{dPVTTC}{dc_i} = \sum_{t=1}^{\infty} \frac{\partial PVTTC}{\partial c_t} \frac{dc_t}{dc_i} \tag{A.2.4}$$

The one-dollar budget increase in year *i* will not alter the budgets for the other years with binding constraints, hence, $dc_t = 0$ for those years. However, there will be changes in spending for the years up to and including year *m* with non-binding constraints, and also for years after year *m* when there are no constraints. The years summed in equation A.2.4 fall into four groups.

- $t \le m$ and $\lambda_t = 0$ (non-binding constraint) for which $\frac{\partial PVTTC}{\partial c_t} = 0$ from equation A.2.1
- $t > m$ (unconstrained) for which $\frac{\partial PVTTC}{\partial c_t} = 0$ from equation A.2.2
- $t \le m$, $\lambda_t > 0$ and $t \ne i$ (binding constraint) for which $dc_t = 0$, and
- $t \le m$, $\lambda_t > 0$ and $t = i$ (binding constraint) for which $\frac{dc_t}{dc_i} = 1$.

Hence, at the constrained optimum, all the terms in equation A.2.4 are zero except that for year *i*, for which $\frac{dc_t}{dc_i} = 1$, and

$$\lambda_i = -\frac{\partial PVTTC}{\partial c_i} = -\frac{dPVTTC}{dc_i}$$

## A.3 Minimising PVAC subject to maximum roughness and annual budget constraints

The maximum roughness constraint is assumed to be binding. The optimisation problem and Kuhn-Tucker conditions are as follows.

Minimise $PVAC(c_1, c_2, \dots, c_m, c_{m+1}, \dots)$ subject to $R(c_1, c_2, \dots, c_m, c_{m+1}, \dots) = R_{max}$, and $c_t \le B_t$ for all *t = 1* to *m*

The Lagrangian is

$$L = \sum_{t=1}^{\infty} \frac{c_t}{(1+r)^t} + \mu[R(c_1, c_2, \dots, c_m, c_{m+1}, \dots) - R_{max}] + \sum_{t=1}^{m} \lambda_t (c_t - B_t)$$

where $\mu$ is the Lagrange multiplier for the roughness constraint. The conditions for a constrained maximum are

$$\frac{\partial L}{\partial c_t} = \frac{1}{(1+r)^t} + \mu \frac{\partial R}{\partial c_t} + \lambda_t = 0 \text{ for all } t = 1 \text{ to } m \tag{A.3.1}$$

$$\frac{\partial L}{\partial c_t} = \frac{1}{(1+r)^t} + \mu \frac{\partial R}{\partial c_t} = 0 \text{ for all } t = m + 1 \text{ to } \infty \tag{A.3.2}$$

$$\frac{\partial L}{\partial \mu} = R(c_1, c_2, \dots, c_m, c_{m+1}, \dots) - R_{max} = 0 \tag{A.3.3}$$

$$\frac{\partial L}{\partial \lambda_t} = c_t - B_t \le 0 \text{ for all } t = 1 \text{ to } m$$

$$\lambda_t (c_t - B_t) = 0 \text{ for all } t = 1 \text{ to } m$$

$$\lambda_t \ge 0 \text{ for all } t = 1 \text{ to } m$$

$$\mu > 0$$

Equation A.3.2 is multiplied through by $dc_t$, and all instances summed to obtain

$$\sum_{t=m+1}^{\infty} \frac{dc_t}{(1+r)^t} + \mu \sum_{t=m+1}^{\infty} \frac{\partial R}{\partial c_t} dc_t = 0 \qquad \text{(A.3.4)}$$

All instances of equation A.3.1 for which the budget constraint is non-binding, hence $\lambda_t = 0$, can be treated in the same way to obtain a similar expression. Let *N* be the set of years $t < m$ with non-binding budget constraints.

$$\sum_{t \in N} \frac{dc_t}{(1+r)^t} + \mu \sum_{t \in N} \frac{\partial R}{\partial c_t} dc_t = 0 \qquad \text{(A.3.5)}$$

Summing equations A.3.4 and A.3.5 and rearranging

$$\mu = -\frac{\sum_{t \in N} \frac{dc_t}{(1+r)^t} + \sum_{t=m+1}^{\infty} \frac{dc_t}{(1+r)^t}}{\sum_{t \in N} \frac{\partial R}{\partial c_t} dc_t + \sum_{t=m+1}^{\infty} \frac{\partial R}{\partial c_t} dc_t} \qquad \text{(A.3.6)}$$

Substituting equation A.3.6 into the instance of equation A.3.1 for which $t = i$, and multiplying it by $dc_i$,

$$\frac{dc_i}{(1+r)^i} - \left[\frac{\sum_{t \in N} \frac{dc_t}{(1+r)^t} + \sum_{t=m+1}^{\infty} \frac{dc_t}{(1+r)^t}}{\sum_{t \in N} \frac{\partial R}{\partial c_t} dc_t + \sum_{t=m+1}^{\infty} \frac{\partial R}{\partial c_t} dc_t}\right] \frac{\partial R}{\partial c_i} dc_i + \lambda_i dc_i = 0 \qquad \text{(A.3.7)}$$

Along the maximum roughness frontier where equation A.3.3 holds,

$$dR = \sum_{t=1}^{\infty} \frac{\partial R}{\partial c_t} dc_t = 0$$

This can be split into two parts

$$dR = \sum_{t=1}^{m} \frac{\partial R}{\partial c_t} dc_t + \sum_{t=m+1}^{\infty} \frac{\partial R}{\partial c_t} dc_t = 0 \qquad \text{(A.3.8)}$$

For years $t < m$ with binding budget constraints, a small change in the constraint for year *i* will have no effect on spending other than for year *i* itself, so $dc_t = 0$. Spending will only change for years $t < m$ with non-binding constraints, that is, for the years $t \in N$. and for year *i* itself. So equation A.3.8 can be rewritten as

$$dR = \frac{\partial R}{\partial c_i} dc_i + \sum_{t \in N} \frac{\partial R}{\partial c_t} dc_t + \sum_{t=m+1}^{\infty} \frac{\partial R}{\partial c_t} dc_t = 0$$

which gives

$$-\frac{\partial R}{\partial c_i} dc_i = \sum_{t \in N} \frac{\partial R}{\partial c_t} dc_t + \sum_{t=m+1}^{\infty} \frac{\partial R}{\partial c_t} dc_t \qquad \text{(A.3.9)}$$

Substituting equation A.3.9 into equation A.3.7 causes a cancellation of terms in equation A.3.7 leaving

$$\frac{dc_i}{(1+r)^i} + \sum_{t \in N} \frac{dc_t}{(1+r)^t} + \sum_{t=m+1}^{\infty} \frac{dc_t}{(1+r)^t} + \lambda_i dc_i = 0$$

which can be rewritten as

$$dPVAC + \lambda_i dc_i = 0$$

$$\lambda_i = -\frac{dPVAC}{dc_i}$$

# Appendix B – Computer methods for selected parts of the BITRE road maintenance model

This appendix is to assist people wishing to develop maintenance optimisation models similar to the one developed for this report. It explains and provides computer code for enumerating all possible treatment options subject to a minimum time interval between treatments and testing for dominated options. It also addresses strategies and methods to reduce computer run times.

Explanations of Mathematica commands in the code shown are available from Wolfram Language and System Documentation: https://reference.wolfram.com/language/ref/menuitem/DocumentationCenter.html

## B.1 Setting up treatment options

This section sets out the computer code for deriving treatment timing and type combinations for an analysis period of 40 years and minimum time interval between treatments of eight years.

The code finds all the 7837 possible combinations of treatment timings and with up to five combinations of three treatment types ($3^5$ = 243) starting from {1, 9, 17, 25, 33} (five treatments; the first in year one with eight years between them) to { } (no treatments at all over the 40 years). The number of possible combinations of treatment types is less than 243 where there are fewer than five treatments. For example, with two treatments, say in years 7 and 23, there a $3^2$ = 9 possible combinations of treatment types.

If the two key parameters, a 40-year analysis period and 8-year minimum time interval between treatments, are altered, changes must be made to the programming to accommodate a different value for the maximum number of treatments that can be carried out during the analysis period. For example, reducing the analysis period to 30 years while retaining the 8-year minimum treatment time interval, reduces the maximum number of treatments from five to four.

Two methods are shown. The first is Visual Basic code for an Excel macro using nested loops. The second is Mathematica code using functional programming.

### B.1.1 Visual Basic code — nested loops

In the Visual Basic Macro using nested loops, the list of five treatment years is filled from left to right. For each value, *y*, in the list in positions 1 to 4, the next value on the right (positions 2 to 5) is iterated from y + 8 to 41. When the value in a position reaches 41, the loop is exited.

The code creates two arrays

- iTrmtYears: An array of five integer values for the years which up to five treatments occur eg. {1, 9, 17, 25, 33}. For timing combinations with less than five treatments, the blanks are filled with 41 eg. {3, 21, 41, 41, 41} means treatments in years 3 and 21 only.
- iTypes: An array of five integer values for the types of treatments: 1 = resurface; 2 = resurface with shape correction; 3 = rehabilitation, for example {1, 1, 2, 3, 1}. For timing combinations with less than five treatments, the blanks are filled with 1s, for example {1, 3, 1, 1, 1} where only two treatments are carried out during the analysis period.

```
Dim iTrmtYears(1 To 5) As Integer
Dim iTypes(1 To 5) As Integer
Dim iPeriod As Integer 'The length of the analysis period plus one
Dim iMinterval As Integer 'The minimum number of years between treatments
Sub TreatmentTimes()
iPeriod = 41
iMinterval = 8
```

```
iTrmtYears(1) = 1 'The first element of the treatment times array is set
to one.
Do Until iTrmtYears(1) > iPeriod
   iTrmtYears(2) = Application.Min(iTrmtYears(1) + iMinterval, iPeriod)
   Do Until iTrmtYears(2) > iPeriod
      iTrmtYears(3) = Application.Min(iTrmtYears(2) + iMinterval, iPeriod)
      Do Until iTrmtYears(3) > iPeriod
         iTrmtYears(4) = Application.Min(iTrmtYears(3) + iMinterval,
         iPeriod)
         Do Until iTrmtYears(4) > iPeriod
            iTrmtYears(5) = Application.Min(iTrmtYears(4) + iMinterval,
            iPeriod)
            Do Until iTrmtYears(5) > iPeriod
               Range("Treatment_description").Range(Cells(1, 1),
               Cells(1, 5)).Value = iTrmtYears 'enters the array of five
               treatment times into the spreadsheet starting in the cell
               given the range name "Treatment_description"
               TreatmentTypes 'calls the sub-routine below
               iTrmtYears(5) = iTrmtYears(5) + 1
            Loop 'Visual Basic command to end a Do loop
            iTrmtYears(4) = iTrmtYears(4) + 1
         Loop
         iTrmtYears(3) = iTrmtYears(3) + 1
      Loop
      iTrmtYears(2) = iTrmtYears(2) + 1
   Loop
   iTrmtYears(1) = iTrmtYears(1) + 1
Loop
End Sub
```

The following sub-routine works through all possible combinations of three treatment types for the number of treatment years. If the number of treatments during the analysis period is *n*, there will be $3^n$ combinations of treatment types, ranging from 3 for one treatment to 243 for five treatments. Treatment types are assessed in descending order from the most expensive treatment, rehabilitation (code 3), to the least expensive, resurface (code 1). The reason for going in descending order is that if a treatment time–type combination consisting of all rehabilitations is rejected on the grounds it does not meet the technical restrictions (for example, the maximum permitted roughness), then it will be impossible for any treatment type combinations in which rehabilitations are replaced with less expensive treatments in the same years to satisfy the technical restrictions. Processing time is saved by not assessing any further treatment type combinations in those years with less expensive treatments (codes 1 or 2).

```
Sub TreatmentTypes()
iTypes(1)= SetType(1) 'See below for the SetType function.
While iTypes(1) > 0
   iTypes(2)= SetType(2)
   While iTypes(2) > 0
      iTypes(3)= SetType(3)
      While iTypes(3) > 0
         iTypes(4)= SetType(4)
```

```
        While iTypes(4) > 0
          iTypes(5)= SetType(5)
          While iTypes(5) > 0
            Range("Treatment_description").Range(Cells(1,6),
            Cells(1,10)).Value = iTypes 'enters the list of five
            treatment types into the spreadsheet immediately to the right
            of the list of five treatment times
            Calculate 'Runs the spreadsheet model to project pavement
            condition forward and estimate road agency and user costs.
            Note that the 'Calculate' command is required when automatic
            calculation has been switched off to reduce unnecessary
            processing. Further code at this point of the program, not
            shown here, records the model results that have to be
            retained, storing them elsewhere in the spreadsheet.
            iTypes(5)= iTypes(5)-1
          Wend 'Visual Basic command to end a While loop
          iTypes(4)= iTypes(4)-1
        Wend
        iTypes(3)= iTypes(3)-1
      Wend
      iTypes(2)= iTypes(2)-1
    Wend
    iTypes(1)= iTypes(1)-1
Wend
End Sub

Function SetType(i) 'If the treatment year for a given position in
iTrmtYear is 41, it is a blank and a 1 is entered for the treatment type
in the same position in iType.
If iTrmtYears(i) = iPeriod Then
   SetType = 1
Else
   SetType = 3
EndIf
End Function
```

### B.1.2 Mathematica code — functional programming

In the Mathematica code shown below, key words are coloured dark blue and user-defined functions dark teal. The symbol '/@' (short for the Mathematica command 'Map') creates a list of outputs from the function on the left of the symbol with each member of the list to the right inputted to the function. Functions used only once can be defined as 'pure functions' with input variables to a function represented by '#' or '#1', '#2' etc. with '&' at the end of the function separating it from the inputs. For example, #^2 &[3] gives 9 and (#1 + #2) &[3, 4] gives 7.

The Mathematica code to derive the treatment time and type combinations consists of just three statements.

```
fNextTrmtYears[y_]:=
Select[Join[Most[y],Last[y]+Range[1,40,8]],LessThan[41]];
```

```
YearsList =
NestWhileList[fNextTrmtYears,Range[1,40,8],!EqualTo[#][{40}]&];

TrmtTypeCombos = Tuples[{3,2,1},#]&/@Range[5];
```

The first two lines of code produce a list of 7836 lists of treatment year combinations as follows: *YearsList* = {{1, 9, 17, 25, 33}, {1, 9, 17, 25, 34}, {1, 9, 17, 25, 35}, …, {1, 9, 17, 25, 40}, {1, 9, 17, 25}, …, {32, 39}, {32, 40}, {32}, {33}, {34}, {35}, {36}, {37}, {38}, {39}, {40}}.

Unlike the Visual Basic program, blanks are not required where the number of treatments is less than five. The code needs only to be run once at the start of the program and held in memory to reuse for each segment processed. The null treatment option, { }, was omitted because leaving the pavement untreated for 40 years invariably violates the technical restrictions. For short analysis periods, for example 20 years, it was necessary to include the null treatment option.

The first statement declares the function fNextTrmtYears[y_] and second statement applies the function recursively. NestWhileList[*f, expression, test*] generates a list of the results of applying the function *f* repeatedly, starting with the *expression*, and continuing until the *test* no longer yields the result 'True'.

The second statement sets the first element in *YearsList* to Range[1, 40, 8] = {1, 9, 17, 25, 33}. Then it expands the list by applying the function *fNextTrmtYears* to the last element in the list to obtain the next element. It keeps doing this while the last element is not equal to {40}. The process is halted just before the null combination is reached. To include the null element in *YearsList*, the expression `!EqualTo[#][{40}]&` must be replaced with `!EqualTo[#][{}]&`.

The function fNextTrmtYears[y_], when applied to a list of treatment years, say the third element in *YearsList*, *y* = {1, 9, 17, 25, 35}

- adds the last element (*Last[y]*), 35, to *Range*[1, 40, 8] = {1, 9, 17, 25, 33} to obtain {36, 42, 50, 58, 66}
- joins it to *y*, after dropping the last element of *y* (*Most[y]*), to obtain {1, 9, 17, 25, 34, 42, 50, 58, 66}
- then selects all values less than 41 to obtain the next member of *YearsList*, {1, 9, 17, 25, 34}.

Applying *fNextTrmtYears* to the list y = {7, 20, 40} transforms it thus

- Last[y]+Range[1,40,8]] = 40 + {1, 9, 17, 25, 33} = {41, 49, 57, 65, 73}
- Join[Most[y], previous result] = Join[{7, 20}, {41, 49, 57, 65, 73}] = {7, 20, 41, 49, 57, 65, 73}
- Select[previous result, LessThan[41]] = Select[{7, 20, 41, 49, 57, 65, 73}, LessThan[41]] = {7, 20}

The combinations in *YearsList* are in the same order as produced by the Visual Basic code above.

The third statement in the code sets up a list of five lists, *TrmtTypeCombos*, each containing all possible combinations of the elements {3, 2, 1}. For one treatment, the combinations are {{3}, {2}, {1}}. For two treatments, the combinations are {{3, 3}, {3, 2}, {3, 1}, {2, 3}, {2, 2}, {2, 1}, {1, 3}, {1, 2}, {1, 1}}, and so on up to five treatment types. This is done by applying Mathematica's 'Tuples' function to the elements {3,2,1} to obtain the combinations for each length from 1 to 5, that is, over Range[5] = {1, 2, 3, 4, 5}.

For each combination of treatment times, the maintenance implications are assessed for each possible combination of treatment types. For example, for the treatment time combination {4, 19, 36}, all $3^3$ = 27 type combinations of three would be assessed. As with the Visual Basic code, the treatment types are written out in descending order so that, if the combination with all rehabilitations fails to meet the technical restrictions, 'throw' and 'catch' functions can be employed to exit the function to avoid unnecessary testing for combinations with less effective treatment types.

## B.2 Testing for dominated options

This section discusses stage 2 of the optimisation process for annual budget constraints in Chapter 6 in which the list of up to 581,485 options for a segment was reduced by eliminating 'dominated' options, defined as options that could never appear in any constrained optimum solution.

The process whereby the list of non-dominated options is developed from the list of all options that are technically feasible produced in stage 1 was described in Section 6.3. In brief, the non-dominated options list starts with the first member of in the list of all options as the seed. The non-dominated options list is revised and expanded by taking each option in turn from the all-options list, starting from the second option, — call it option *A* — and comparing it with each option in the non-dominated list, option B. If option *A* is found to dominate any option in the non-dominated list, the latter is removed. If option *A* is not dominated by any option in the non-dominated list, it is added to the list. The process continues until the end of the all-options list is reached.

In the computer model, the dominance test applied was, with budget constraints for the first *m* years, option *A* dominates option *B* if,

$$min\left[(PVTTC^B - PVTTC^A), (c_1^B - c_1^A), \left(c_2^B - c_2^A\right), \cdots, (c_m^B - c_m^A)\right] \geq 0$$

No two options will be identical so all *m + 1* elements in the expression will never equal zero. If option *A* has a lower PVTTC value than option *B*, then $PVTTC^B - PVTTC^A > 0$. For *A* to dominate *B*, there is a further requirement that agency spending in each budget constrained year under option A, not be lower than under option B. If spending under option B in, say, year 5 was lower than under option A, then $c_5^B - c_5^A < 0$, the above test expression would be negative and the test would fail. It would be concluded that option A does not dominate option B. In the unlikely event that the PVTTC values were equal, option A would dominate option B if no years had lower spending under option B than under option A, which implies that option *A* was less costly in at least one year.

In the case of the simple numerical example in Table 6.1, the left-hand expression of the dominance test would give zero values for segments 1, 2, 3 where option A dominates, and –140,000 for segment 4 where neither option dominates.

Having tested whether option A dominates option B, it is then necessary to apply the reverse test to determine whether option B dominates option A.

$$min\left[(PVTTC^A - PVTTC^B), (c_1^A - c_1^B), \left(c_2^A - c_2^B\right), \cdots, (c_m^A - c_m^B)\right] \geq 0$$

In the case of the simple numerical example in Table 6.1, the left-hand expression of the reverse test would give negative values for all four segments.

If both tests fail, then neither option can be eliminated from further consideration by the model. Both have to be included in the non-dominated list. Either one could be subsequently removed if found to be dominated in a subsequent test.

The process of building up the non-dominated list was undertaken using Mathematica's Fold command, Fold[f, {a, b, c, d}] = f[ f[ f[a, b], c], d].

```
NonDominatedList = Fold[function to compare options and build non-
dominated list, {First[full list of options]}, Rest[full list of options]]
```

The 'function to compare options and build non-dominated list' returned the non-dominated list, modified by adding the option from the full options list if dominant with dominated options deleted or not comparable, or unchanged if the option from the full options list was dominated by any options in the non-dominated list.

## B.3 Ways to reduce computer run times

Ways to reduce processing time by the model included

- Calculating the lists of treatment time and type options and the list of discount factors at the commencement of the program, then retaining them in memory for the whole model run.
- For each segment, making projections of vehicle numbers and calculating constants that depend on segment data before commencing to evaluate treatment options for the segment, then retaining the projections and constants in memory until all options for the segment had been evaluated.

- Projecting cracking over the analysis period only once for each set of treatment times. Since all three treatment types reset surface age to zero, the projected cracking is identical regardless of treatment types.
- Using 'throw' and 'catch' functions to immediately cease processing an option when it was found to violate a technical constraint. Such cases occurred mostly for options with long intervals of time without treatment during which pavement condition deteriorated severely.
- Where a segment violated a technical constraint at the start of the analysis period, such as having a roughness above 6.3 m/km IRI, a treatment was mandatory in year one. There was no need to assess options that do not have a treatment in year one.
- For each list of treatment times, the first list of treatment types to be evaluated was all rehabilitation treatments . If this violated a technical constraint, it was not necessary to evaluate further combinations of treatment types with less effective treatments because they would inevitably violate the technical constraints as well.
- The no-treatments option {{ },{ }} was not evaluated because it was not technically feasible for a 40-year analysis period. A pavement neglected for that that length time would exceed the maximum allowable roughness level within that period. The no-treatments option had to be reinstated for sensitivity tests involving shorter analysis periods.

# Appendix C – Depreciation calculation

Depreciation was added to agency costs at the end of the last year of the analysis period with the aim of approximating the PVTTC from the end of the analysis period to infinity.

Ideally, the model would be invariant to implementing a treatment at the end of year 40 and not implementing it because the depreciation value changes by an offsetting amount. This was achieved approximately for two of the three treatment types. The maximum depreciation amount was set at the rehabilitation cost with the design pavement strength for year 40 and a roughness of 5.2 m/km IRI. Part of this amount, the cost of resurfacing with cracking at 0.5% (the level just prior to crack initiation) was depreciated linearly with surface age over the years to crack initiation (that is, 12 years for sprayed treatment pavements and 16 years for asphalt mix pavements). Thus, for a pavement just starting to crack in year 40, agency costs were the same whether or not the model resurfaced in year 40. The remainder of the maximum depreciation amount was apportioned linearly according to roughness between the range 1.2 and 5.2 m/km IRI. Pavement age was not used because of the difficulty in predicting the number of years for a new pavement to reach 5.2 m/km IRI.

It was found that this method of calculating depreciation caused a large number of pavements to be given resurfacing with shape correction treatments in the last few years of the analysis period. The reason was that this intermediate treatment, not taken into account in the depreciation schedule, caused a significant saving in depreciation making the treatment too attractive. Typically, the resurfacing with shape correction treatment would be undertaken around a roughness of 3.5 m/km IRI and would reduce roughness by around one IRI unit. A section with a flatter slope was introduced into the depreciation curve over the range 2.5 to 3.5m/km IRI, compensated for by steeper slopes outside the range. The depreciation function is shown in Figure C.1, and the formulas in Table C.1. Depreciation is shown as a function of roughness for a hypothetical one kilometre of road with a sprayed treatment pavement type and design adjusted structural number of 6.93. The depreciation shown is only the component that varies with roughness. The maximum depreciation that varies with roughness ($189,500) is the cost of a rehabilitation at a roughness of 5.2 m/km IRI ($194,200) minus the cost of a resurface ($4,700) with 0.5% cracking. The middle section between 2.5 and 3.5 IRI has the same value as the straight line at the mid-point 3.0 IRI and a slope equal to the cost per IRI unit of roughness improvement from the resurface with shape correction treatment at 0.5% cracking ($26,700) minus the cost of the resurfacing only treatment.

The example here shows that the difficulty in specifying a depreciation schedule that well approximates PVTTC for years beyond the end of the analysis period increases with the number of possible treatment types. Depreciation schedules for maintenance optimisation models may be a worthwhile area of future research. However, if the analysis period extends well beyond the focus period (in our model, a 20-year focus period with a 40-years analysis period), the highly-approximate nature of the depreciation calculation will have only a limited effect on recommended treatments during the focus period. The sensitivity tests reported in Chapter 5 showed how results can be distorted by having a short analysis period shifting greater reliance onto the depreciation function to simulate PVTTC beyond the analysis period.

Figure C.1 Depreciation as a function of roughness

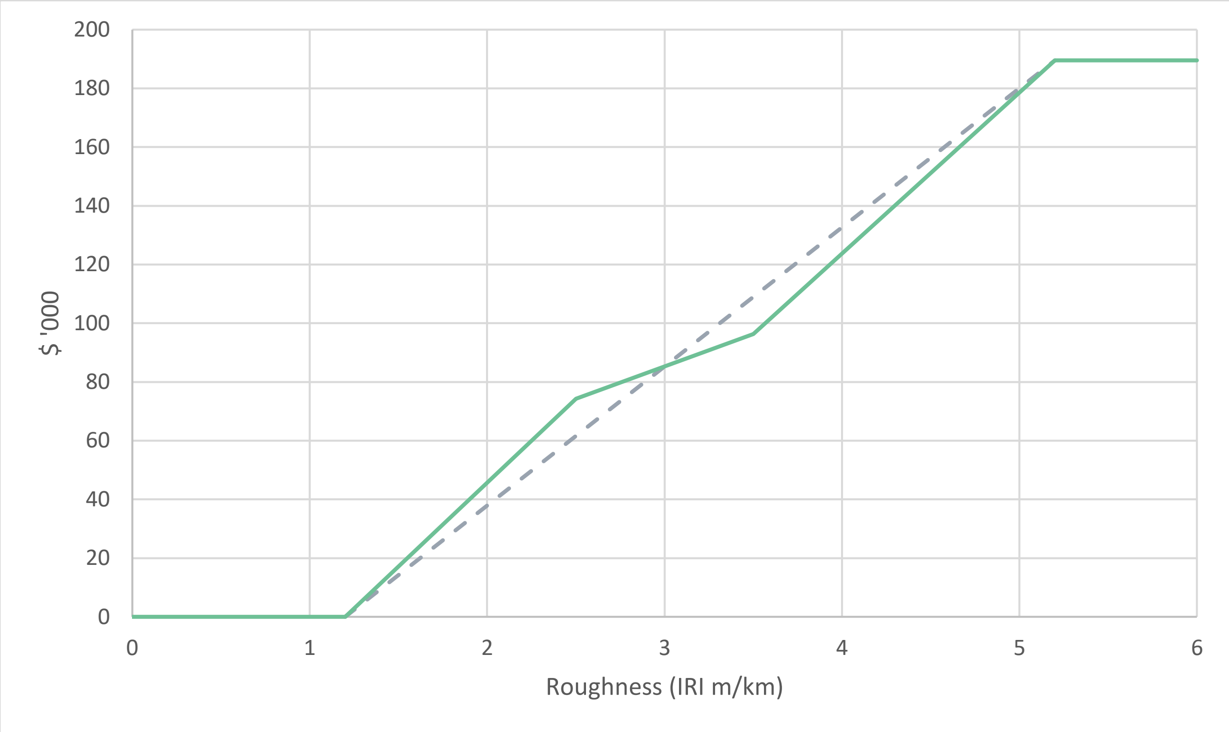



Table C.1 Depreciation formulas

| Roughness range (m/km IRI) | Depreciation formula |
|---|---|
| **0 – 1.2** | 0 |
| **1.2 — 2.5** | (1.8 A – 2 B) (R – 1.2) / 5.5 |
| **2.5 — 3.5** | 0.45 A + (R – 3) B |
| **3.5 – 5.2** | ((10.4 – 2 R) B – (4.64 – 2.2 R) A) / 6.8 |
| **5.2 and above** | A |

where

- *R* = roughness at the end of the analysis period
- *A* = the additional cost of a rehabilitation (treatment 3) at IRI = 5.2, over and above the cost of a resurfacing (treatment 1) with cracking at 0.5% ($189,545 = $194,195 – $4,650 in Figure C.1)
- *B* = the additional cost of resurfacing with shape correction (treatment 2), over and above the cost of resurfacing, both at cracking of 0.5% ($22,000 = $26,650 – $4,650 in Figure C.1)

# Appendix D – ARRB modelling of case study data

ARRB's modelling results for the case study database using HDM-4 are presented here and compared with the BITRE model results to illustrate an alternative modelling approach. ARRB dealt with the 'curse of dimensionality' in three ways. First they aggregated the 2034 road segments into 573 'strategic analysis sections' to enable the analysis to fit within the limits of HDM-4. Second, they employed the method discussed in Chapter 4 of having the model choose between condition-responsive treatment rules instead of times at which particular treatments are implemented. Third, the analysis period was set at 20 years, with a residual value applied at the end of the period.

Five 'treatment alternatives' were specified. Each alternative consisted of a one or more treatment types with condition-based trigger points at which the treatment was implemented. The periodic treatment types were resurfacing, resurfacing with shape correction and rehabilitation. All five alternatives included routine maintenance in the form of patching wide structural cracks and potholes, and edge repair. The treatment alternatives were

- *Base alternative (minimum standard)*: resurface at a target age of 10 years for surface treatment pavements, 15 years for asphalt mix pavements
- *Standard maintenance*: base alternative with resurfacing with shape correction at 3.5 m/km IRI
- *Delayed resurfacing*: standard maintenance with resurfacing delayed to 1.5 times target age
- *Rehabilitation on high distress*: base alternative with rehabilitation at >10% cracking and >10mm mean rut depth
- *Standard Maintenance with rehabilitation at high distress*: base alternative with rehabilitation at 5.4 m/km IRI.

Treatment timing options were tested by allowing commencement of options other than the base alternative to be delayed by varying numbers of years. Each of the four options other than the base alternative could be commenced in any of years 1 to 11 giving rise to 45 options (four alternatives × 11 years and the base alternative) for each segment.

The residual value was calculated as having two components.

- A pavement component set at 90% of the cost of the latest pavement rehabilitation treatment depreciated linearly with roughness between 1.2 and 5.2 m/km IRI, and
- A surfacing component set at the greater of 10% of the cost of the latest pavement rehabilitation treatment and the full cost of the latest surfacing where a surfacing was applied as a separate treatment, depreciated linearly over a life defined as the lesser of the resurfacing target age or time in years until cracking reaches 10%.

HDM-4 projected pavement condition and user costs forward and estimated the net benefits for each alternative for each commencement time compared with the base alternative. The unconstrained optimum for each segment was the option with the lowest of PVTTC out of the 45 possibilities.

Table D.1 summarises the results in terms of percentages of the network (by length) treated, spending for the three periodic treatment types, and in total spending. ARRB also forecast needs for 'other' routine maintenance treatment types (pothole patching, edge repair and crack sealing), but these have not been included in Table D.1 to facilitate comparison with the BITRE model results in identically-formatted Table 5.4. Spending on 'other treatments' was small in relation to the total except in year one when it was $21 million. Over the 20 years, spending on other treatments added up to $41 million, increasing total spending by 4%.

**Table D.1 Summary of ARRB modelling results: unconstrained optimisation**

| Years | Percent of network kilometres treated | | | | Spending ($ millions) | | | |
|---|---|---|---|---|---|---|---|---|
| | Resurf. | RSC | Rehab. | Total | Resurf. | RSC | Rehab. | Total |
| **Totals** | | | | | | | | |
| **1** | 35 | 0 | 1 | 36 | 91 | 3 | 8 | 102 |
| **1 to 10** | 83 | 1 | 28 | 113 | 196 | 9 | 396 | 601 |
| **11 to 20** | 60 | 10 | 8 | 77 | 146 | 66 | 125 | 337 |
| **1 to 20** | 143 | 11 | 36 | 190 | 342 | 75 | 521 | 937 |
| **Annual averages** | | | | | | | | |
| **1 to 10** | 8 | 0 | 3 | 11 | 20 | 1 | 40 | 60 |
| **11 to 20** | 6 | 1 | 1 | 8 | 15 | 7 | 12 | 34 |
| **1 to 20** | 7 | 1 | 2 | 10 | 17 | 4 | 26 | 47 |

Notes: Percentages of network kilometres treated in excess of 100% occur where the same road segments are treated more than once over the time period.
Resurf. = resurface, RSC = resurface with shape correction, Rehab. = rehabilitation

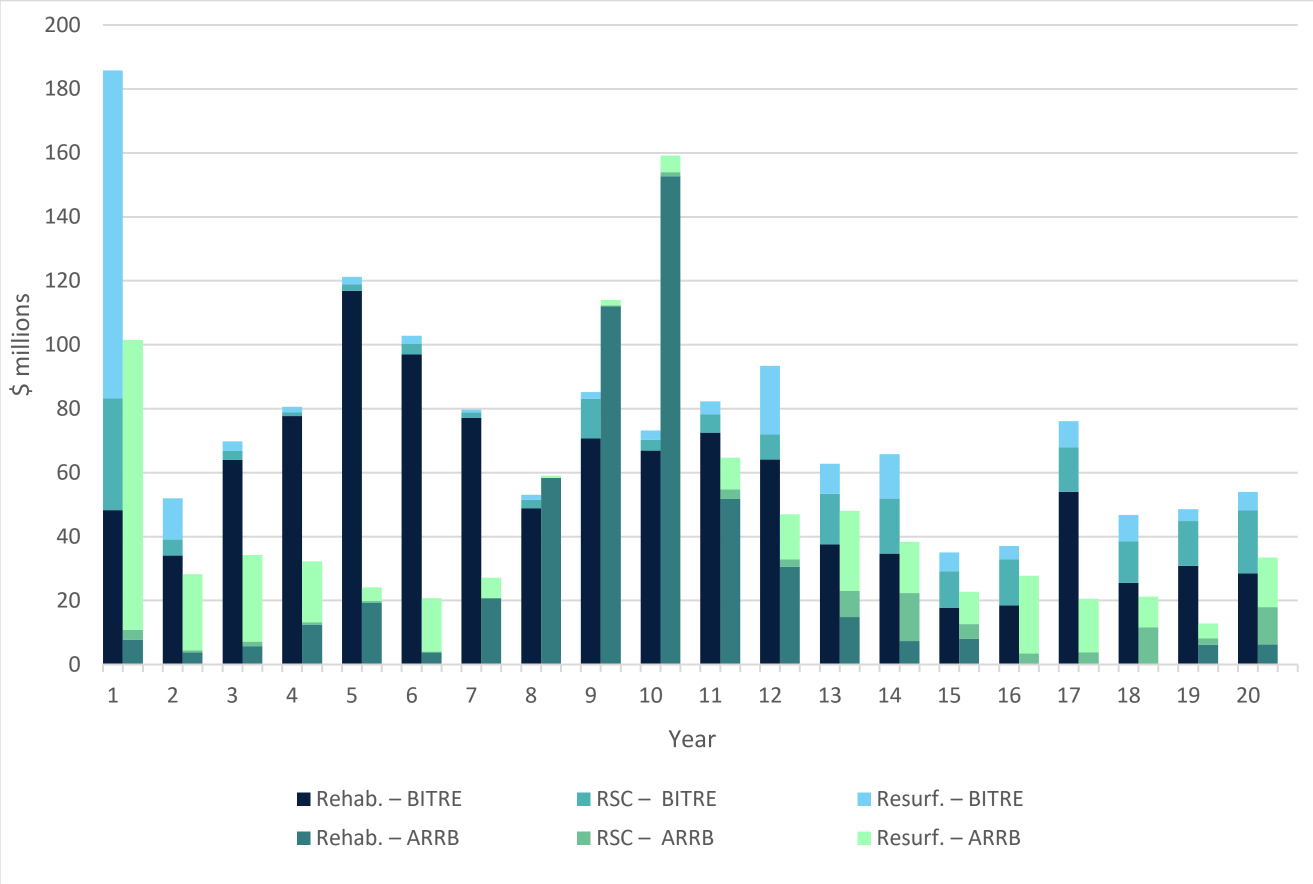
First-year optimal spending was $102 million with 36% of the network by length treated. Most of this was resurfacing — 97% by length and 89% by spending. The split of costs between treatment types for the 20 years was 36% for resurfacing, 8% resurfacing with shape correction and 56% rehabilitation. 'Rehabilitation on high distress' was the optimum alternative for 1,503 kilometres of the network, followed by standard maintenance for 253 kilometres. A total of 713 kilometres or 36% of the network by length was rehabilitated during the 20-year period. Annual average optimal spending over the 20 years was $47 million, with 10% of the network, on average, treated each year. Maintenance activity was more intense during the first 10 years than the second.

Optimisation subject to annual budget constraints was undertaken using the 'ARRB Optimisation Tool', supported by genetic optimisation software 'Evolver'. It was intended to impose annual budget constraints for the first 10 years of 100%, 75%, 50% and 25% of the average year-one to year-10 annual agency costs (including 'other treatments' not shown in Table D.1) of $63.2 million. However, the annual spending levels for the base alternative set the lower limits on annual budget constraints. For the last budget-constrained years, with base alternative spending exceeding $80 million per annum, it was impossible to meet the annual budget constraints, even at the 100% level.

Comparing the ARRB and BITRE results in Tables D.1 and 5.5, over the 20 years, the proportion of the network treated (184% for BITRE and 190% for ARRB) and proportions rehabilitated (37% for BITRE and 36% for ARRB) were remarkably similar. BITRE's model recommended more resurfacing with shape correction and less resurfacing. BITRE's model had 60% higher total spending for the 20 years. The main reason was that the average cost of a rehabilitation treatment was $118 per square metre for the BITRE model compared with $67 for the ARRB model (based on ARRB advice to BITRE received some time after ARRB had completed its modelling).

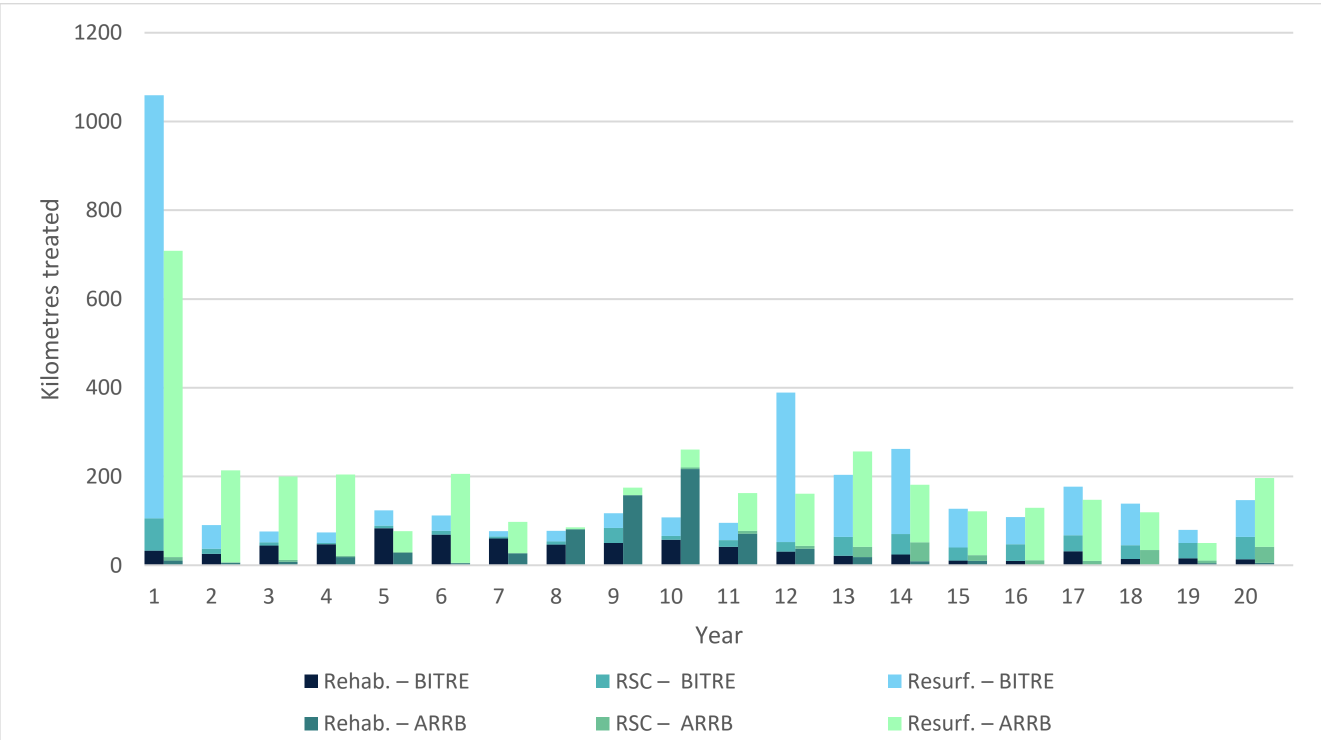
Figures D.1 and D.2 show annual spending and kilometres treated as forecast by the two models. The left bars are the same as in Figures 5.2 and 5.3. The BITRE model had more work done in year one than the ARRB model, 54% compared with 36% of kilometres treated, with the ARRB model pushing resurfacing work into years 2, 3 and 4. This could be a reflection of how the ARRB model considered delaying the commencement of policy options. The BITRE model had rehabilitation work peaking on year five and the ARRB model in years 10 and 17.

**Figure D.1 Forecast optimal expenditure without budget constraints: BITRE and ARRB models**




Notes: For each year, the left bar is the estimate from the BITRE model and right bar from the ARRB model.
Rehab. = rehabilitation, RSC = resurface with shape correction, Resurf. = resurface

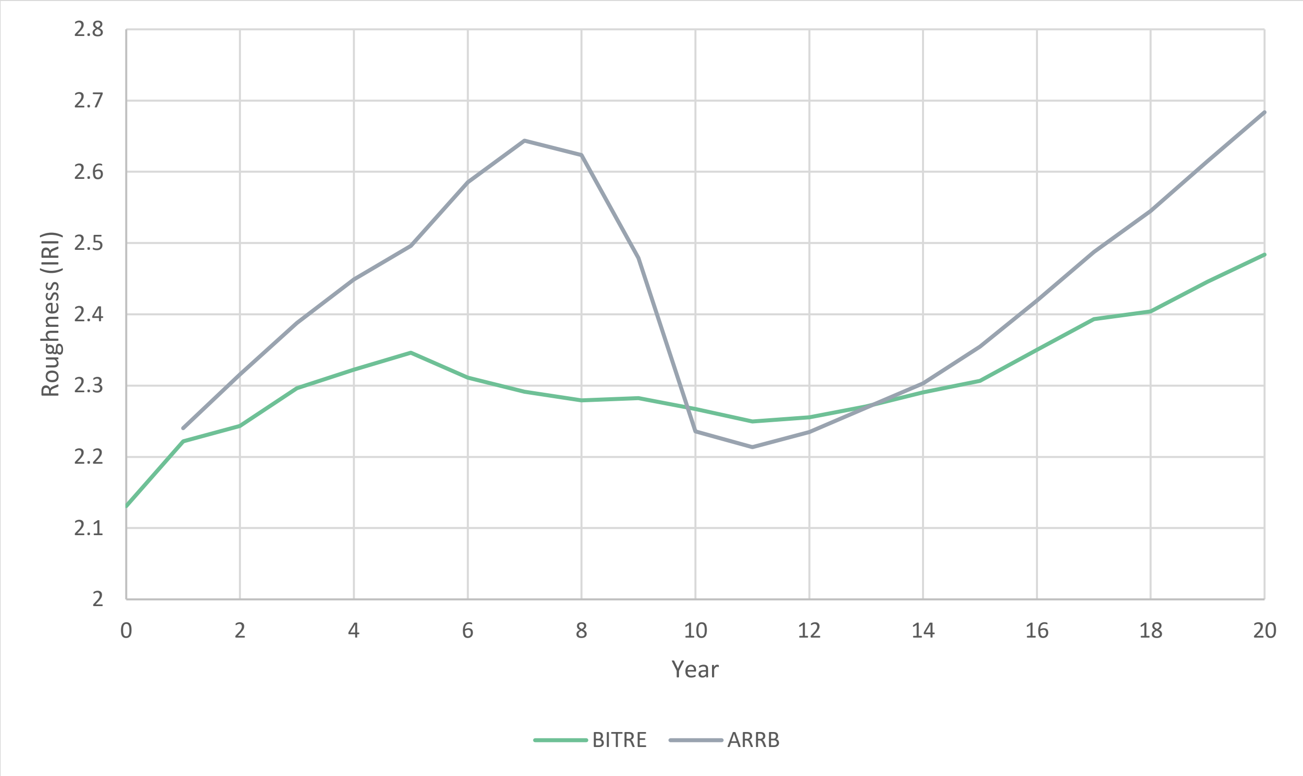
**Figure D.2 Forecast road lengths treated without budget constraints: BITRE and ARRB models**




Note: For each year, the left bar is the estimate from the BITRE model and right bar from the ARRB model.
Rehab. = rehabilitation, RSC = resurface with shape correction, Resurf. = resurface

Figure D.3 shows length-weighted average roughness projected over the 20-year period. The BITRE line is the same as for PVTTC minimisation in Figure 5.8. Both models brought average roughness down in the middle part of the period after the backlog of the network in poor condition was addressed. The ARRB model undertook large amounts of rehabilitation work in years 8 to 12, as evident in Figures D1 and D.2, while the BITRE model spread maintenance works more evenly across years. A possible reason is that the ARRB model had less flexibility to shift treatments between years because it optimised maintenance policies with alternative commencement years instead of individual treatments.

**Figure D.3 Length-weighted average roughness: BITRE and ARRB models**




Note: Starting the vertical axis at 2.0 instead of zero accentuates the vertical difference between the two curves

# References


Abaza, K.A. and Ashur, S.A., 1999. Optimum decision policy for management of pavement maintenance and rehabilitation. *Transportation Research Record,* 1655(1), pp. 8-15.

Abaza, K., Ashur, S., Abu Eisheh, S. and Rabayah, A., 2001. Macroscopic optimum system for the management of pavement rehabilitation. *Journal of Transportation Engineering*, 127(6). pp. 493-500.

Abaza, K., 2002. Optimum flexible pavement life-cycle analysis model. *Journal of Transportation Engineering*, 128(6), pp. 542-549.

Abaza, K., Khatib, I. and Ashur, S., 2004. Integrated pavement management system with a Markovian prediction model. *Journal of Transportation Engineering*, 130(1), pp. 24-33.

Abaza, K.A. and Murad, M.M., 2007. Dynamic probabilistic approach for long-term pavement restoration program with added user cost. *Transportation Research Record*, 1990(1), pp. 48-56.

Abelson, P.W. and Flowerdew, A.D.J., 1975. Models for the economic evaluation of road Maintenance. *Journal of Transport Economics and Policy*, 9(2), pp. 93-114.

Alonso, M. and Yanguas, S., 2001. Analysis of correlations between longitudinal indices in FILTER experiment: Forum of European National Highway Research Laboratories Investigation of longitudinal and transverse evenness of roads experiment. *Transportation Research Record*, 1764(1), pp. 243-253.

Archondo-Callao, R., 2008. *Applying the HDM-4 model to strategic planning of road works*. World Bank, Washington DC.

Australia Transport Assessment and Planning Guidelines (ATAP), 2016. *PV2: Parameter values*. Commonwealth Department of Infrastructure and Regional Development, Canberra.

Australia Transport Assessment and Planning Guidelines (ATAP), 2022. *T2: Cost Benefit Analysis*. Commonwealth Department of Infrastructure, Transport, Regional Development and Communications, Canberra.

Austroads 2007. *Guide to asset management: Part 5B: Roughness*. AGAM05B/07, Austroads, Sydney.

Austroads 2008. *Road Surface Characteristics and Crash Occurrence: A Literature Review*. AP–T96/08, Austroads, Sydney.

Austroads 2010. *Road Safety Engineering Risk Assessment Part 7: Crash Rates Database*. Austroads Technical Report, AP-T152/10, Austroads, Sydney.

Austroads 2018. *Guide to Asset Management — Technical Information Part 15: Technical Supplements*. AGAM15-18, Austroads, Sydney.

Beavis, B. and Dobbs, I., 1990. *Optimisation and stability theory for economic analysis*. Cambridge University Press, Cambridge.

Bekheet, W., Helali, K., Li, Y., Cheetham, A. and Stanciu, D., 2008. Comparison between probabilistic and deterministic pavement management analysis: a case study for Arizona DOT. In *7th International Conference on Managing Pavement Assets* (pp. 23-28).

Bellman, R., 1954. The theory of dynamic programming. *Bulletin of the American Mathematical Society*, 60(6), pp. 503-515.

Bellman R.E., 1961. *Adaptive Control Processes: A guided tour*. Princeton University Press, Princeton, NJ.

Bhandari, A., Harral, C., Holland, E. and Faiz, A., 1987. Technical options for road maintenance in developing countries and the economic consequences. *Transportation Research Record*, 1128, pp. 18-27.

Burningham, S. and Stankevich, N., 2005. *Why road maintenance is important and how to get it done*. Transport note no. TRN-4, World Bank, Washington DC.

Bureau of Infrastructure, Transport and Regional Economics (BITRE), 2021. *Yearbook 2021: Australian Infrastructure and Transport Statistics*. Statistical Report, BITRE, Canberra ACT.

Butt, A.A., Shahin, M.Y., Carpenter, S.H. and Carnahan, J.V., 1994. Application of Markov process to pavement management systems at network level. In *3rd international conference on managing pavements* (Vol. 2, pp. 159-172). Washington, DC: National Research Council, Transportation Research Board.

Büttler, H.J. and Shortreed, J.H., 1978. Investment planning of a road link. *Transportation Research*, *12*(5), pp. 357-366.

Carnahan, J.V., Davis, W.J., Shahin, M.Y., Keane, P.L. and Wu, M.I., 1987. Optimal maintenance decisions for pavement management. *Journal of Transportation Engineering*, 113(5), pp. 554-572.

Carnahan, J.V., 1988. Analytical framework for optimizing pavement maintenance. *Journal of Transportation Engineering, 114*(3), pp. 307-322.

Cary, W. N.; Irick, P. E., 1960. *The pavement serviceability-performance concept*. (PDF). Transportation Research Board.

Cenek P.D, Davies R.B., Henderson R.J., 2012. *Crash risk relationships for improved road safety management*. New Zealand Transport Agency research report 488.

Chan, C.Y., Huang, B., Yan, X. and Richards, S., 2009. Relationship between highway pavement condition, crash frequency, and crash type. *Journal of Transportation Safety & Security*, 1(4), pp.268-281.

Chan, W.T., Fwa, T.F. and Tan, C.Y., 1994. Road-maintenance planning using genetic algorithms. I: Formulation. *Journal of transportation engineering*, 120(5), pp. 693-709.

Chan, W.T., Fwa, T.F. and Hoque, K.Z., 2001. Constraint handling methods in pavement maintenance programming. *Transportation Research Part C: Emerging Technologies*, *9*(3), pp. 175-190.

Chen, L., Henning, T.F., Raith, A. and Shamseldin, A.Y., 2015. Multiobjective optimization for maintenance decision making in infrastructure asset management. *Journal of Management in Engineering*, 31(6), 04015015.

Cooper, D. R. C., P. G. Jordan, and J. C. Young., 1980. *The effect on traffic speeds of resurfacing a road*. Transport and Road Research Laboratory, Crowthorne, Berkshire, UK.

Cornes, R., 1992. *Duality and modern economics*. Cambridge University Press, Cambridge.

Davis, C.F. and Van Dine, C.P., 1988. *Linear programming model for pavement management. Transportation Research Record,* 1200, pp. 71-75.

de la Garza, J.M., Akyildiz, S., Bish, D.R. and Krueger, D.A., 2011. Network-level optimization of pavement maintenance renewal strategies. *Advanced Engineering Informatics*, 25(4), pp. 699-712.

Deb, K., 1999. Multi-objective genetic algorithms: Problem difficulties and construction of test problems. *Evolutionary computation*, 7(3), pp. 205-230.

Dekker, R., 1996. Applications of maintenance optimization models: a review and analysis. *Reliability engineering & system safety*, 51(3), pp. 229-240.

Department of Infrastructure and Regional Development: Australian Government (DITRD), 2013. *Notes on Administration for Land Transport Infrastructure Projects: 2014-15 to 2018-19*. DITRD, Canberra.

Donnges, C., Edmonds, G. and Johannessen, B., 2007. *Rural road maintenance: Sustaining the benefits of improved access*. International Labour Office, Bangkok.

Durango, P.L. and Madanat, S.M., 2002. Optimal maintenance and repair policies in infrastructure management under uncertain facility deterioration rates: an adaptive control approach. *Transportation Research Part A: Policy and Practice*, 36(9), pp. 763-778.

Durango-Cohen, P.L., 2007. A time series analysis framework for transportation infrastructure management. *Transportation Research Part B: Methodological*, 41(5), pp. 493-505.

Feighan, K.J., Shahin, M.Y., Sinha, K.C. and White, T.D., 1988. Application of dynamic programming and other mathematical techniques to pavement management systems. *Transportation Research Record*, 1200, pp. 90-98.

Feighan, K.J., Shahin, M.Y., Sinha, K.C. and White, T.D., 1989. A prioritization scheme for the micro PAVER pavement management system. *Transportation Research Record*, 1215, pp. 89-100.

Ferreira, A., Antunes, A. and Picado-Santos, L.I.S., 2002a. Probabilistic segment-linked pavement management optimization model. *Journal of Transportation Engineering*, 128(6), pp. 568-577.

Ferreira, A., Picado-Santos, L. and Antunes, A., 2002b. A segment-linked optimization model for deterministic pavement management systems. *International Journal of Pavement Engineering*, 3(2), pp. 95-105.

Ferreira A. L. and Queiroz C., 2012. A deterministic section-linked optimization model for road network maintenance management. *3rd International Conference on Engineering Optimization*, Rio de Janeiro, Brazil 1-5 July.

Friesz, T.L. and Fernandez, J.E., 1979. A model of optimal transport maintenance with demand responsiveness. *Transportation Research Part B: Methodological*, 13(4), pp. 317-339.

Fwa, T.F., Chan, W.T. and Tan, C.Y., 1994a. Optimal programming by genetic algorithms for pavement management. *Transportation Research Record*, 1455, pp. 31-40.

Fwa, T.F., Tan, C.Y. and Chan, W.T., 1994b. Road-maintenance planning using genetic algorithms. II: Analysis. *Journal of transportation engineering*, 120(5), pp. 710-722.

Fwa, T.F., Chan, W.T. and Tan, C.Y., 1996. Genetic-algorithm programming of road maintenance and rehabilitation. *Journal of Transportation Engineering*, 122(3), pp. 246-253.

Fwa, T.F., Chan, W.T. and Hoque, K.Z., 1998. Analysis of pavement management activities programming by genetic algorithms. *Transportation Research Record*, 1643(1), pp. 1-6.

Fwa, T.F., Sinha, K.C. and Riverson, J.D., 1988. Highway routine maintenance programming at network level. *Journal of Transportation Engineering*, 114(5), pp. 539-554.

Fwa, T.F., Chan, W.T. and Hoque, K.Z., 2000. Multiobjective optimization for pavement maintenance programming. *Journal of transportation engineering*, 126(5), pp. 367-374.

Gao, L. and Zhang, Z., 2008. Robust optimization for managing pavement maintenance and rehabilitation. *Transportation Research Record*, 2084(1), pp. 55-61.

Gao, L., Xie, C., Zhang, Z. and Waller, S.T., 2012. Network-level road pavement maintenance and rehabilitation scheduling for optimal performance improvement and budget utilization. *Computer-Aided Civil and Infrastructure Engineering*, 27(4), pp. 276-287.

Gerchak, Y. and Waters, W.G., 1978. Models for the economic evaluation of road maintenance: a comment. *Journal of Transport Economics and Policy*, 12(1), pp. 103-108.

Golabi, K., Kulkarni, R.B. and Way, G.B., 1982. A statewide pavement management system. *Interfaces*, 12(6), pp. 5-21.

Goldberg E.D., 1989. *Genetic algorithms in search, optimisation and machine learning.* Addison-Wesley Publishing Company, Menlo Park, California.

Golroo, A. and Tighe, S.L., 2012. Optimum genetic algorithm structure selection in pavement management. *Asian Journal of Applied Sciences*, *5*(6), pp. 327-341.

Grivas, D.A., Ravirala, V. and Schultz, B.C., 1993. State increment optimization methodology for network-level pavement management. *Transportation Research Record*, 1397, pp. 25-33.

Gu, W., Ouyang, Y. and Madanat, S., 2012. Joint optimization of pavement maintenance and resurfacing planning. *Transportation Research Part B: Methodological*, *46*(4), pp. 511-519.

Hajek, J.J. and Phang, W.A., 1988. Prioritization and optimization of pavement preservation treatments. *Transportation Research Record*, 1216, pp. 58-68.

Harvey, M.O., 2012. *Optimising road maintenance*. Discussion paper no. 2012-12, International Transport Forum, Paris.

Harvey, M.O., 2015. Commercial road supply with incentive regulation. *International Journal of Sustainable Transportation*, 9(4), pp. 241-253.

Heggie, I.G., 1995. *Management and financing of roads: An agenda for reform*. World Bank technical paper no. 275, Washington DC, USA.

Heggie I.G. and Vickers P., 1998. *Commercial management and financing of roads*. World Bank technical paper no. 409, Washington DC, USA.

Holland J.H., 1975. *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor, Michigan.

Infrastructure Australia (IA) 2016. *Instructure plan: Priorities and reforms for our nation's future.* Infrastructure Australia, Sydney.

Infrastructure Australia (IA) 2019. *An Assessment of Australia's Future Infrastructure Needs: The Australian Infrastructure Audit 2019.* Infrastructure Australia, Sydney.

Jawad, D. and Ozbay, K., 2006. Probabilistic lifecycle cost optimization for pavement management at the project-level. In *Transportation Research Board, 2006 Annual Meeting CD-ROM*.

Jorge, D. and Ferreira, A., 2012. Road network pavement maintenance optimisation using the HDM-4 pavement performance prediction models. *International Journal of Pavement Engineering*, 13(1), pp. 39-51.

Justo-Silva, R., Ferreira, A. and Flintsch, G., 2021. Review on machine learning techniques for developing pavement performance prediction models. *Sustainability*, 13(9), p. 5248.

Kalembo C., Jeihani M., Saka S., 2012. *Evaluation of the Impact of Pavement Roughness on Vehicle Gas Emissions in Baltimore County*. Transport Research Board Annual Meeting

Khan, M.U., Mesbah, M., Ferreira, L. and Williams, D.J., 2014. Development of road deterioration models incorporating flooding for optimum maintenance and rehabilitation strategies. *Road & Transport Research: A Journal of Australian and New Zealand Research and Practice*, 23(1), pp. 3-24.

Konak, A., Coit, D.W. and Smith, A.E., 2006. Multi-objective optimization using genetic algorithms: a tutorial. *Reliability Engineering & System Safety*, 91(9), pp. 992-1007.

Kulkarni, R.B. and Miller, R.W., 2003. Pavement management systems: Past, present, and future. *Transportation Research Record*, 1853(1), pp. 65-71.

Lee, J., Abdel-Aty, M. and Nyame-Baafi, E., 2020. Investigating the effects of pavement roughness on freeway safety using data from five states. *Transportation Research Record*, 2674(2), pp. 127-134.

Li, N., Xie, W.C. and Haas, R., 1996. Reliability-based processing of Markov chains for modeling pavement network deterioration. *Transportation Research Record*, 1524(1), pp. 203-213.

Li, N., Haas, R. and Xie, W.C., 1997a. Investigation of relationship between deterministic and probabilistic prediction models in pavement management. *Transportation Research Record*, 1592(1), pp. 70-79.

Li, N., Huot, M. and Haas, R., 1997b. Cost-effectiveness-based priority programming of standardized pavement maintenance. *Transportation Research Record*, 1592(1), pp. 8-16.

Li, N., Haas, R. and Huot, M., 1998. Integer programming of maintenance and rehabilitation treatments for pavement networks. *Transportation Research Record*, 1629(1), pp. 242-248.

Li, Y. and Madanat, S., 2002. A steady-state solution for the optimal pavement resurfacing problem. *Transportation Research Part A: Policy and Practice*, 36(6), pp. 525-535.

Liu, C., Hammad, A. and Itoh, Y., 1997. Maintenance strategy optimization of bridge decks using genetic algorithm. *Journal of transportation engineering*, 123(2), pp. 91-100.

Mahoney, J.P., Ahmed, N.U. and Lytton, R.L., 1978. Optimization of pavement rehabilitation and maintenance by use of integer programming. *Transportation Research Record*, 674, pp. 15-22.

Mamlouk, M., Vinayakamurthy, M., Underwood, B.S. and Kaloush, K.E., 2018. Effects of the international roughness index and rut depth on crash rates. *Transportation research record*, 2672(40), pp.418-429.

Markow, M. and Balta, W., 1985. Optimal rehabilitation frequencies for highway pavements. *Transportation Research Record*, *1035*, pp. 31-43.

Martin, T. and Choummanivong, L., 2018. *Predicting the performance of Australia's arterial and sealed local roads.* ARR390. Australian Road Research Board. Vermont South, Victoria.

Mbwana, J.R. and Turnquist, M.A., 1996. Optimization modeling for enhanced network-level pavement management system. *Transportation Research Record*, *1524*(1), pp. 76-85.

McLean J. and Foley G., 1998. *Road surface characteristics and condition: effects on road users.* ARRB Transport Research Ltd, Research report ARR314, Vermont South, Victoria.

Medury, A. and Madanat, S., 2013. Incorporating network considerations into pavement management systems: A case for approximate dynamic programming. *Transportation Research Part C: Emerging Technologies*, *33*, pp. 134-150.

Medury, A. and Madanat, S., 2014. Simultaneous network optimization approach for pavement management systems. *Journal of Infrastructure Systems*, 20(3), pp. 04014010-1-7.

Meneses, S. and Ferreira, A., 2013. Pavement maintenance programming considering two objectives: maintenance costs and user costs. *International Journal of Pavement Engineering*, 14(2), pp. 206-221.

Morcous, G. and Lounis, Z., 2005. Maintenance optimization of infrastructure networks using genetic algorithms. *Automation in Construction*, 14(1), pp. 129-142.

Morosiuk G., Riley, M and Odoki J.B., 2004. *Modelling road deterioration and works effects, Highway development and management series.* Volume 6. World Road Association (PIARC).

Naudé C., Tsolakis D., and Toole T., 2008. An "accelerated" approach to funding road rehabilitation. *23rd ARRB Conference*, Adelaide, Australia.

Naudé C. and Toole T., 2012. Evaluation of the accelerated road rehabilitation program (ARRP). *25th ARRB Conference*, Perth, Australia.

Newbery, D.M., 1988. Road damage externalities and road user charges. *Econometrica: Journal of the Econometric Society*, 56(2), pp. 295-316.

Ng, M., Zhang, Z. and Waller, S.T., 2011. The price of uncertainty in pavement infrastructure management planning: An integer programming approach. *Transportation Research Part C: Emerging Technologies*, 19(6), pp. 1326-1338.

Odoki, J.B. and Kerali, G.R., 2006. *Analytical framework and model descriptions. Highway development and management series*. Volume 4. World Road Association (PIARC).

Opus 1999. *Review of VOC-pavement roughness relationships contained in Transfund's project evaluation manual*. Opus International Consultants Ltd, Central Laboratories Report 529277.00, Lower Hutt, New Zealand.

Ortiz-García, J.J., Costello, S.B. and Snaith, M.S., 2006. Derivation of transition probability matrices for pavement deterioration modeling. *Journal of transportation engineering*, 132(2), pp. 141-161.

Ouyang, Y. and Madanat, S., 2004. Optimal scheduling of rehabilitation activities for multiple pavement facilities: exact and approximate solutions. *Transportation Research Part A: Policy and Practice*, 38(5), pp. 347-365.

Ouyang, Y. and Madanat, S., 2006. An analytical solution for the finite-horizon pavement resurfacing planning problem. *Transportation Research Part B: Methodological*, 40(9), pp. 767-778.

Parkman C.C., Bradbury T., Peeling D. and Booth C., 2012. *Economic, Environmental and Social Impacts of Changes in Maintenance Spend on Local Roads in Scotland*. Transport Research Laboratory, Report prepared for Transport Scotland.

Paterson W. and Watanatada T., 1985. *Measuring road roughness and its effects on user cost and comfort. In: Relationships between vehicle speed, ride quality, and road roughness.* Philadelphia: American Society for Testing and Materials, 89–110.

Paterson, W., 1986. International roughness index: Relationship to other measures of roughness and riding quality. *Transportation Research Record*, 1084, pp 49-59.

Paterson, W.D.O., 1987. *Road deterioration and maintenance effects: Models for planning and management*. Published for the World Bank by The John Hopkins University Press, Baltimore.

Paterson, W. and Attoh-Okine, B., 1992. Summary models of paved road deterioration based on HDM-III. *Transportation Research Record*, 1344, pp. 99-105.

Pilson, C., Hudson, W.R. and Anderson, V., 1999. Multi-objective optimization in pavement management by using genetic algorithms and efficient surfaces. *Transportation Research Record*, 1655(1), pp. 42-48.

Qiao, Y., Dawson, A., Parry, T. and Flintsch, G.W., 2018. Immediate effects of some corrective maintenance interventions on flexible pavements. *International Journal of Pavement Engineering*, 19(6), pp.502-508.

Rashid, M.M. and Tsunokawa, K., 2012. Trend curve optimal control model for optimizing pavement maintenance strategies consisting of various treatments. *Computer-Aided Civil and Infrastructure Engineering*, *27*(3), pp. 155-169.

Ravirala, V. and Grivas, D.A., 1994. Nonpreemptive goal programming methodology for developing annual pavement program. *Transportation Research Record*, 1455, pp. 50-57.

Ravirala, V. and Grivas, D.A., 1995. Goal-programming methodology for integrating pavement and bridge programs. *Journal of Transportation Engineering*, 121(4), pp. 345-351.

Ravirala, V., Grivas, D.A., Arminio Jr, S. and Garrabrant, R., 1997. Goal-oriented optimization procedure for long-term highway investment planning. *Transportation Research Record*, 1592(1), pp. 17-25.

Roth, G. (ed.), 2006. *Street smart: competition, entrepreneurship and the future of roads*. Transaction Publishers, New Brunswick.

Sathaye, N. and Madanat, S., 2011. A bottom-up solution for the multi-facility optimal pavement resurfacing problem. *Transportation Research Part B: Methodological*, 45(7), pp. 1004-1017.

Scheinberg, T. and Anastasopoulos, P.C., 2010. Pavement preservation programming: a multi-year multi-constraint optimization methodology. In *presentation at the 89th Annual Meeting of the Transportation Research Board and publication in the Transportation Research Record*.

Shah, Y.U., Jain, S.S. and Parida, M., 2014. Evaluation of prioritization methods for effective pavement maintenance of urban roads. *International Journal of Pavement Engineering*, 15(3), pp. 238-250.

Shtayat, A., Moridpour, S., Best, B. and Rumi, S., 2022. An overview of pavement degradation prediction models. *Journal of Advanced Transportation*, 2022, pp. 1-15.

Small, K.A., Winston, C. and Evans, C.A., 2012. *Road work: A new highway pricing and investment policy*. Brookings Institution Press, Washington DC.

Tack, J.N. and J. Chou, E.Y., 2002. Multiyear pavement repair scheduling optimization by preconstrained genetic algorithm. *Transportation Research Record*, 1816(1), pp. 3-8.

Tehrani, S., Cowe Falls, L. and Mesher, D., 2017. Effects of pavement condition on roadway safety in the province of Alberta. *Journal of Transportation Safety & Security*, 9(3), pp.259-272.

Toole T., Michel N., Walker B. and Booth I., 2006. Long term road performance estimates and treatment strategies in Tasmania. *22nd ARRB Conference*, Canberra, Australia.

Toole, T. and Roper, R., 2014. *Case study of the maintenance needs of the non-urban road corridors of the National Land Transport Network*. Contract report prepared by ARRB for BITRE.

Torres-Machí, C., Chamorro, A., Videla, C., Pellicer, E. and Yepes, V., 2014. An iterative approach for the optimization of pavement maintenance management at the network level. *The Scientific World Journal*, 2014.

Tsunokawa, K. and Schofer, J.L., 1994. Trend curve optimal control model for highway pavement maintenance: Case study and evaluation. *Transportation Research Part A: Policy and Practice*, 28(2), pp. 151-166.

Tsunokawa, K. and Ul-Islam, R., 2003. Optimal pavement design and maintenance strategy for developing countries: An analysis using HDM-4. *International Journal of Pavement Engineering*, 4(4), pp. 193-208.

Tsunokawa, K., Van Hiep, D. and Ul-Islam, R., 2006. True optimization of pavement maintenance options with what-if models. *Computer-Aided Civil and Infrastructure Engineering*, 21(3), pp. 193-204.

Ul-Islam, R. and Tsunokawa, K., 2006. A combined optimization-simulation methodology for pavement maintenance. In *Research into Practice: 22nd ARRB Conference ARRB*, Canberra, Australia.

Wang, F., Zhang, Z. and Machemehl, R.B., 2003. Decision-making problem for managing pavement maintenance and rehabilitation projects. *Transportation Research Record*, 1853(1), pp. 21-28.

Wang, K.C., Nguyen, V. and Zaniewski, J.P., 2007. Genetic algorithms–based network optimization system with multiple objectives. *Transportation Research Record*, 2016(1), pp. 85-95.

Watanatada T., Harral C., Paterson W.D.O., Dhareshwar, A.M., Bhandari A., Tsunokawa K., 1987. *The Highway Design and Maintenance Standards Model: Volume 1, Description of the HDM-III Model*. Published for the World Bank by The John Hopkins University Press, Baltimore, Maryland, USA.

World Road Association (PIARC) 2014. *The importance of road maintenance*. PIARC, France.

Wu, Z. and Flintsch, G.W., 2009. Pavement preservation optimization considering multiple objectives and budget variability. *Journal of Transportation Engineering*, 135(5), pp. 305-315.

Yeniay, Ö., 2005. Penalty function methods for constrained optimization with genetic algorithms. *Mathematical and computational Applications*, 10(1), pp. 45-56.

Yeo, H., Yoon, Y. and Madanat, S., 2013. Algorithms for bottom-up maintenance optimisation for heterogeneous infrastructure systems. *Structure and Infrastructure Engineering*, *9*(4), pp. 317-328.

Yepes, V., Torres-Machi, C., Chamorro, A. and Pellicer, E., 2016. Optimal pavement maintenance programs based on a hybrid greedy randomized adaptive search procedure algorithm. *Journal of Civil Engineering and Management*, 22(4), pp. 540-550.

Yu, B. and Lu, Q., 2014. Empirical model of roughness effect on vehicle speed. *International Journal of Pavement Engineering*, 15(4), pp. 345-351.

Zeitlow G.J., 2006. Role of the private sector in managing and maintaining roads. In: Roth G (ed.), *Street Smart*. Transaction Publishers, New Brunswick, pp. 347–371.

Zimmerman, K.A., 1995. *Pavement management methodologies to select projects and recommend preservation treatments*. NCHRP Synthesis 222, Transportation Research Board.

Connecting
Australia

www.bitre.gov.au