Minnesota Department of Transportation

RESEARCH SERVICES

Office of Policy Analysis, Research & Innovation

Optimal Contract Mechanism Design for Performance-Based Contracts

Diwakar Gupta, Principal Investigator
Department of Mechanical Engineering
Industrial and Systems Engineering Program
University of Minnesota

August 2011

Research Project
Final Report 2011-18

LRRB
LOCAL ROAD RESEARCH BOARD

Your Destination...Our Priority

All agencies, departments, divisions and units that develop, use and/or purchase written materials for distribution to the public must ensure that each document contain a statement indicating that the information is available in alternative formats to individuals with disabilities upon request. Include the following statement on each document that is distributed:

To request this document in an alternative format, call Bruce Lattu at 651-366-4718 or 1-800-657-3774 (Greater Minnesota); 711 or 1-800-627-3529 (Minnesota Relay). You may also send an e-mail to bruce.lattu@state.mn.us. (Please request at least one week in advance).

Technical Report Documentation Page

1. Report No. MN/RC 2011-18	2.	3. Recipients Accession No.
4. Title and Subtitle Optimal Contract Mechanism Design for Performance-Based Contracts		5. Report Date August 2011 6.
7. Author(s) Diwakar Gupta, Aditya Vedantam, Justin Azadivar		8. Performing Organization Report No.
9. Performing Organization Name and Address Department of Mechanical Engineering Industrial and Systems Engineering Program University of Minnesota 111 Church St SE Minneapolis, MN 55455		10. Project/Task/Work Unit No. CTS Project #2010041 11. Contract (C) or Grant (G) No. (c) 89261 (wo) 173
12. Sponsoring Organization Name and Address Minnesota Department of Transportation Research Services Section 395 John Ireland Boulevard, MS 330 St. Paul, MN 55155		13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code
15. Supplementary Notes http://www.lrrb.org/pdf/201118.pdf		

16. Abstract (Limit: 250 words)

This report explores the possibility of implementing a performance-based contract (PBC) in highway construction and maintenance. PBC allows a state transportation agency (STA) to use an incentive structure to induce the contractor to take a long-term interest in the functionality of the highway without excessive cost to the STA. First, an overview of contract mechanisms and existing PBCs in industry is provided. Then, a summary of performance-based specifications illustrates that contractors can adopt construction and maintenance policies to significantly improve pavement performance. Finally, a framework for choosing PBCs to achieve STAs' goals is provided, with a discussion of some of the difficulties associated with implementation. It can be seen that, in a competitive bidding process, the additional cost of incentives is not larger than the additional value gained by higher-quality work.

17. Document Analysis/Descriptors Performance-based contracting, Contracting, Highway construction, Road construction		18. Availability Statement No restrictions. Document available from: National Technical Information Services, Alexandria, Virginia 22312	
19. Security Class (this report) Unclassified	20. Security Class (this page) Unclassified	21. No. of Pages 53	22. Price

Optimal Contract Mechanism Design for Performance-Based Contracts

Final Report

Prepared by:

Diwakar Gupta
Aditya Vedantam
Justin Azadivar

Department of Mechanical Engineering
Industrial and Systems Engineering Program
University of Minnesota

August 2011

Published by:

Minnesota Department of Transportation
Research Services Section
395 John Ireland Boulevard, Mail Stop 330
St. Paul, Minnesota 55155

Acknowledgments

Researchers are grateful to the Minnesota Department of Transportation's (MnDOT) Richard B. Kjonaas, deputy state aid engineer, who served as technical liaison for this project, and Dan Warzala, research project coordinator, who served as administrative liaison. Both helped researchers obtain important feedback on a variety of technical issues during the execution of this project. The researchers are also grateful to members of the technical advisory panel (TAP), in particular MnDOT's Maureen Jensen, Tom Ravn, and Curt Turgeon, who provided their valuable advice through a series of TAP meetings during the course of this project.

Contents

List of Tables

List of Figures

Executive Summary

Nearly all transportation-related pavement construction work is performed by contractors. State transportation agencies (STAs) have used a variety of mechanisms like design-build, warranty and build-maintain contracts to ensure constructed pavement's quality and service life. More recently Performance-Based Specifications (PBS) have been used to ensure high levels of as-built quality by paying contractors based on predicted life-cycle cost.

The objective of a contract mechanism should be to ensure that the built structure maintains adequate functionality throughout its service life and that this objective is realized in a cost-effective manner. In this report, we propose Performance-Based Contracts (PBCs), which view construction contracts as long-term agreements between STAs and contractors. PBCs can be used by STAs to provide incentives or disincentives to contractors based on in-use performance of the pavement. Contracting on measurable performance and providing the right incentives has the potential to align contractor interests with that of the STAs and lead to better-quality roads.

In this report, we describe different types of contracts and provide examples of the use of PBCs in a variety of industries. We study Closed Form Solutions (CFS) and establish factors that contractors can control to affect product performance. We find that contractors, by adopting the right construction methods and performing timely preemptive maintenance, can significantly improve product quality. We provide an inventory of performance metrics that STAs can use to evaluate contractor performance, which will form the basis of a performance-based contract. Finally, we propose a framework for choosing parameters of PBCs to achieve STAs' goals, while making a note of some of the implementation issues associated with PBCs.

1 Introduction

This report contains a discussion of the tools available and the considerations needed to approach implementing a more complex model for designing procurement contracts for transporation construction projects.

We begin in chapter 2 by providing a brief overview of the different types of procurement contracts used in industry starting from simple supply contracts to more complex Performance-Based Contracts (PBCs). Contract mechanism design can be viewed as a principal-agent problem in which the principal pays the agent to perform a service. Contracts serve to align the interests of the agent with those of the principal, while the latter pursues its self interest. We propose PBCs as an effective means to achieve lower lifecycle costs in the transportation construction industry by giving contractors incentive to perform preemptive maintenance. PBCs utilize a greater number of contract parameters than traditional contracts. We note that while PBCs are better able to affect contractors' decisions, they do so at the cost of greater implementation effort. In Chapter 3, examples of PBC implementation in the healthcare, rail maintenance, defense, computing and transportation industries are discussed. Their benefits and implementation issues are also highlighted.

Chapter 4 discusses payment schemes used in Performance-Based Specifications (PBS) mechanisms that use Closed Form Solutions (CFS) to predict certain distress types, e.g., rutting and fatigue cracking. It is noted that CFS require shorter computational time relative to distress prediction done using Mechanistic Empirical Pavement Design Guide (MEPDG). Studying the CFS reveals the set of parameters that contractors can control in the design mix. Contractors can also control the timing and extent of preventive maintenance and make efforts to reduce the variability in construction practices, both of which affect pavement performance. Chapter 4 lists all parameters that contractors can control to affect the amount of observed distress and gives examples of actions that have the potential to increase the usable life of pavements, and a set of performance metrics that can be used to evaluate contractor performance.

Chapter 5 uses the insights gained by studying the PBS model to build a mathematical framework for implementing PBCs. It demonstrates analytical results showing the ability of the state transportation agencies (STAs) to use incentives to control maintenance policies of contractors under the conditions of a PBC. Doing so does not increase the overall cost to the STA or to the contractor beyond the additional performance quality gained, but rather aligns their incentives so that STAs goals are achieved while the contractor seeks to maximize its profit. It then discusses some implementation issues particular to PBCs that must be considered.

2 Types of Contracts

The transportation construction industry needs mechanisms for letting and managing construction projects through which optimal product performance can be realized throughout it's service life. Designing an optimal contract requires understanding existing contracts not only in the transportation industry, but also in other industries. This section describes examples of various contracts used in industry, in increasing order of complexity.

2.1 Supply Contracts

In a typical supply contract, a buyer must decide how much of a material or service to order from a seller to satisfy an exogenous demand, often in the presence of storage (or obsolescence) and shortage costs. If the buyer orders too much it will be subject to storage/obsolescence costs, and if it orders too little it will incur shortage costs. The seller in turn maximizes its profit by incentivizing the buyer to order more. If the buyer and the seller are concerned only with maximizing their individual profits without coordinating with the other party, then that self-serving focus leads to sub-optimal performance of the entire supply chain. By using coordinating contracts both players can increase their individual profits and maximize the total supply chain profit. Typical contract parameters are wholesale price, timing of deliveries, quantity of purchase in each period, provision for returns by the buyer, revenue sharing, and side payments [1]. However these contracts include no stipulations on the performance of the product after delivery.

A major source of uncertainty in this setting is demand. Furthermore, disruptions in supply (natural disasters), uncertain yield, and lead time variations can add to supply uncertainty and may be accounted for in contract design.

2.2 Product Warranty Contracts

A warranty is a manufacturer's assurance to a buyer that a product or service is or shall be as represented. Warranties that are provided for consumer durables (e.g., household appliances, cars) and industrial and commercial products (e.g., equipment used in hospitals, parts of an aircraft) are called product warranties [2]. These are offered by manufacturers to signal the product quality and provide a guarantee against failure of the product for a prespecified period of time. Thus, warranty contracts add an extra dimension to supply contracts - a guarantee of product quality. If the product is found to be defective during the warranty period, the manufacturer is liable to repair or replace the product.

For example, a company might offer a one year limited warranty (limited to the terms set) against defects in material and workmanship whereby it will either (1) exchange the product for a new one, or (2) repair the product at no charge, or (3) refund the purchase price of the product, or else (4) repair the product at a prorated cost. The contract parameters here would be the cost of the product (which includes warranty cost), duration and type of warranty (e.g., free replacement, prorated warranty) and terms specifying what constitutes performance on such contracts (e.g. repair within 7 working days). Customers can also purchase add-on options to extend warranty period, change performance location (e.g. onsite versus depot), or get coverage over items/components excluded from standard terms.

The purpose of warranties is twofold - protection and promotion. Warranty terms that specify the extent of coverage in the event of mishandling or misuse protect the manufacturer from lawsuits, while the consumer is protected from purchasing a defective product. Product warranties are also used as a promotional strategy by manufacturers, especially when marketing new products.

2.3 Procurement Warranty Contracts

In simple transactions involving consumer or commercial goods, a government agency buying such products will be treated in the same way as any other customer purchasing the product with warranty. However, certain government acquisitions (e.g., a new fleet of tanks or fighter jets or the construction of a new bridge) which are typically characterized by a high degree of uncertainty in the product development process, need additional assurance from the seller. Procurement warranty contracts are often used in such settings. They require the seller or the manufacturer to guarantee a specified level of performance by meeting the minimum performance requirement or reliability goals of the buyer.

Procurement warranties in which the contractor's objective is to meet design and performance requirements are called Essential Performance Requirements (EPR). Here the contractor is required to ensure that the number of failures do not exceed a pre-agreed amount. Alternatively, the contractor may be given a fixed upfront payment, based on the average cost per repair multiplied by the expected number of failures, and held responsible for keeping the product operational. In these cases, the contractor is motivated to keep failures at a minimum to increase its profits. Such types of procurement warranties, where an incentive is offered to the seller to increase the reliability of the items after they are put into service, are called Reliability Improvement Warranties (RIW) (see [3]). However, the reliability of new products for which little historical data is available is difficult to predict, which makes it difficult for the buyers to assess risk. As we will see later, PBCs reduce this uncertainty by contracting based on in-use quality.

Procurement warranties act as a risk hedging tool by shifting some of the development and acquisition risk from the buyer to the contractor. The contractor also benefits by having an opportunity to make extra profit through incentives. When procurement is driven by a low-cost sealed-bid mechanism, the presence of warranties gives a competitive advantage to those contractors who build higher-quality products at competitive prices. In the transportation construction industry, Wisconsin Department of Transportation (WisDOT) began using pavement warranties in 1995 and is one of the first states to explore this contracting method. Today, thirty-five states have used some form of warranty provision on construction projects. In [4], a study to compare the cost and performance of different types of warranted and non-warranted pavements is reported. The study, which was conducted with almost 12 years of performance data, concluded that warranted pavements have lower distress levels and better ride quality than similar pavements that were not constructed under warranty. The reason proposed for better performance is the contractor's greater ownership over material and construction methods. Median costs for warranted hot mix asphalt (HMA) pavements over flexible base were also lower than for corresponding non-warranted HMA pavements. Possible reasons put forward to explain lower costs are that mix designs and paving operations may be more cost-effective when the contractor has control over design and production processes. Whereas procurement warranty contracts have been widely used in the construction industry, the results from their implementation are not uniformly positive. For example, a study conducted by the Illinois DOT observed that even with warranties, contractors did not use innovative methods or give more attention to detail during construction [5]. Section 2.7 highlights how the proposed PBCs improve upon procurement warranty contracts.

2.4 Service Contracts

Often self-maintenance of equipment can be costly and time consuming. In some cases consumers may not be well-equipped to perform maintenance work (e.g., when maintenance requires special training or when customers do not have customized hoists and tools to perform maintenance activity). In such cases the manufacturer or a third party can offer a service or "extended warranty" contract, whereby maintenance activities on the equipment are outsourced for a stipulated period of time.

A service contract includes cost of the service, duration and terms of maintenance (e.g., type of work covered). Service contracts cover two types of maintenance activities - corrective and preventive. In the pavement construction industry, preventive maintenance is referred to as pavement preservation and corrective actions are called pavement rehabilitation. In corrective maintenance, unplanned repairs on the equipment are performed after the equipment breaks down whereas in preventive maintenance, actions are taken to prevent future defects and increase the reliability of the system.

2.5 Performance-Based Specifications

STAs use PBS in the procurement of construction services. The STA specifies design specs that a structure to be built by the contractor (e.g., pavement, bridge, or building) must meet. The design specifications correspond to a certain predicted performance and a service life expectation. After the contractor completes construction, the predicted performance and service life are calculated again, based on built quality. The contractor is then paid a lump sum at project completion that includes an incentive if the anticipated as-built lifecycle cost is less than the anticipated as-designed lifecycle cost, and a disincentive otherwise. In Section 4 we discuss PBS methodology in greater detail.

STAs use PBS to mitigate the risk of the contractor building the product using inferior materials and construction methods. The Incentive-Disincentive (I/D) scheme ensures that the contractor uses a mix and construction processes that are expected to perform at least as well as the design, leading to a lower expected total lifecycle cost. In Figure 2.1, expected performance for the design mix is compared with the predicted performance for the as-built mix. The amount of incentive that the contractor is paid is proportional to the area between the two curves. The key contract parameters are the design specifications, I/D scheme used and the performance metrics on which the contractor is rewarded or penalized.

Hence, PBS ensure that the as-built specs are as good as the as-design specs specified. However this alone does not guarantee a good performance of the pavement throughout its life. Contracts are based on performance projections as they relate to inputs. The contractor therefore attempts to match the as-built specs to the design specs, which does not necessarily imply that the actual performance of the structure will be as good or better than the expected performance. The contractor's responsibility ends with the completion of the structure and it is not directly concerned with either actual performance or preventive maintenance. Further, the contractor is paid based on performance prediction methods that have limited accuracy. In the next section we propose PBCs

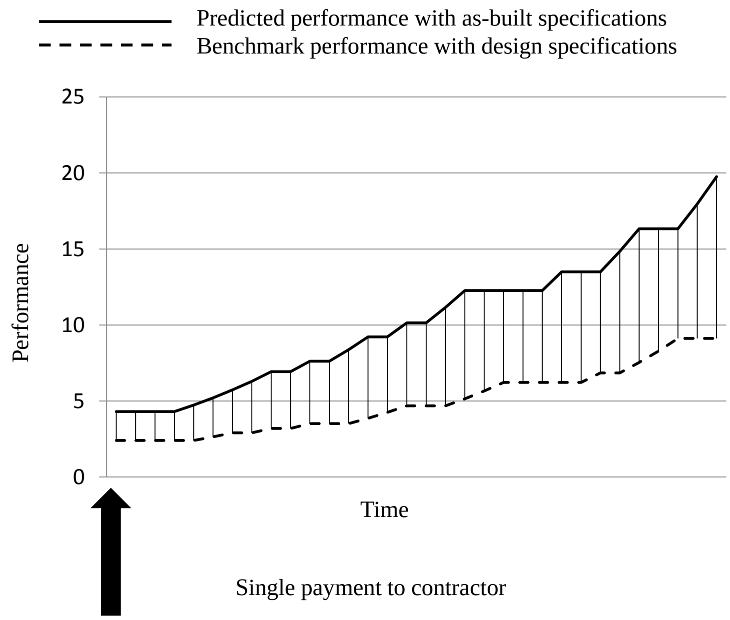

Figure 2.1: Performance-Based Specifications

that overcome many of the shortcomings of PBS, by paying the contractor based on in-use performance and shifting the focus from measuring inputs to measuring realized performance, which is what matters to road users.

2.6 Performance-Based Contracts

In PBCs, the STA announces a set of design specifications that the contractor tries to meet and that are used for developing performance metrics, i.e., for predicting expected product performance over a set of agreed upon performance metrics. The contractor is free to choose materials, components, construction methods, and preventive maintenance activities. The contractor is paid its bid amount upon completion (subject to meeting some minimum standards) and recurring payments are made throughout the structure's life, based on performance measurements taken at regular intervals. The contractor is given incentives if the in-use performance is better than the calculated performance benchmarks and disincentives otherwise.

In Figure 2.2, we show the magnitude and frequency of payments for a hypothetical example. The major contract parameters are the I/D scheme (i.e., the amount and timing of payments), the frequency of measurements, and the contracted performance criteria.

By paying contractors incentives/disincentives for measured performance instead of the quality of inputs, the STA can encourage the contractor to use construction methods and mix quality that improves performance of the structure throughout its life. The contractor is given flexibility in choosing materials and construction methods, which fosters innovation. Further, the contractor is encouraged to take timely preventive maintenance actions, which eliminates the need to perform costly corrective maintenance in the future. PBCs also reduce the impact of errors associated with

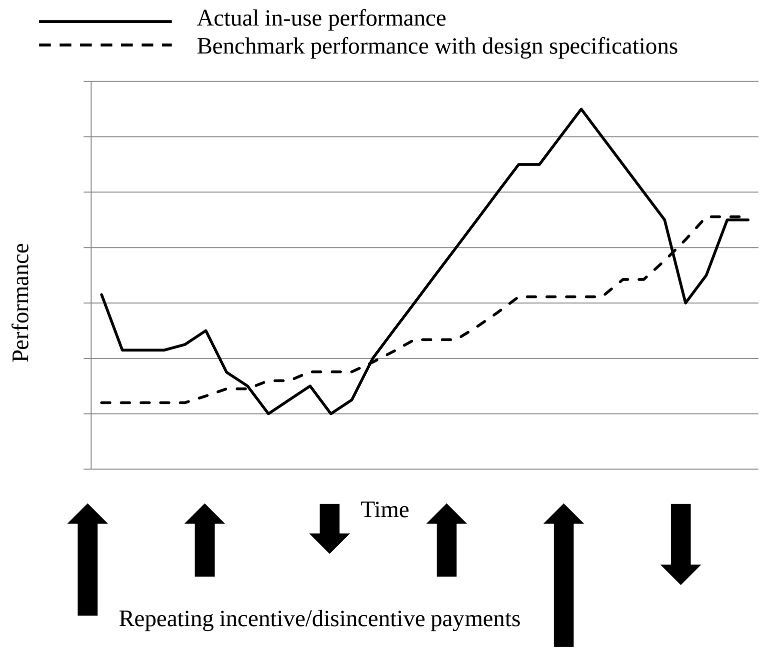

Figure 2.2: Performance-Based Contracts

performance prediction. While PBS contracts predict both as-built and expected performance, PBCs are based on agreed upon performance benchmarks, which need not make use of prediction models. The predicted as-built performance in PBS is replaced by actual measured performance in PBC, making I/D payments more accurate.

Whereas warranty contracts in construction require contractors to perform corrective maintenance if the performance is below a minimum threshold, PBCs encourage preventive maintenance. Further, warranty contracts only require the contractor to meet the bare minimum standards whereas PBCs encourage contractors to improve performance over the contracted life of the product. With PBCs, we predict that STAs can reduce costs of maintaining structures, which are higher when maintenance work is corrective rather than preventive. Thus PBCs have the potential to improve overall product performance and reduce total life cycle cost. Table 2.1 highlights the major differences between PBS and PBC approaches. In Table 2.2, we highlight the major differences between PBS and warranty contracts.

PBC design needs to take into account several factors. In-use performance of a structure can be affected by external factors, such as temperature values that lie outside the range of design specifications or unexpected changes in traffic loading. Because such externalities are uncertain and beyond a contractor's control, the contractor should not be penalized for the resulting product performance, which may require STAs to specify a mechanism for adjusting performance benchmarks. Thus, PBCs create an environment in which STAs and contractors share responsibility for total life cycle cost, while improving in-use performance by incentivizing the contractor to perform preventive maintenance on the structure. Below we highlight the key takeaways from this section.

2.7 Key Takeaways

Traditional procurement contracts are simple to administer because they have few contract parameters. PBCs have a greater number of contract parameters and emphasize performance, but they are relatively harder to implement. In theory, PBCs are superior to traditional contracts because they offer more accurate and targeted performance incentives, leading to superior product quality and lower total life cycle cost.

Table 2.1: Differences between PBS and PBC

	Performance-Based Specifications (PBS)	**Performance-Based Contracts (PBC)**
1	Contractor incentivized based on predicted performance of as-built product	Incentives/Disincentives based on in-use quality **(more accurate)**
2	Lump sum payment to contractor at project completion	Payment of bid amount at project completion and recurring payments as incentives or disincentives over structure's life. **(fosters innovation and preventive maintenance)**
3	As-built specifications must be as good as design specifications	Contractor is given flexibility in choosing materials and construction methods. Emphasis is on long-term performance (output) rather than quality of inputs.
4	Contractor responsibility ends with construction of structure	Contractor–STA interaction continues throughout the contracted life of the structure

Table 2.2: Differences between PBC and Warranty Contracts

	Warranty Contracts	**Performance-Based Contracts (PBC)**
1	Require contractor to perform corrective maintenance when performance falls below a threshold during the warranty period	Incentives are provided to the contractor to perform preventive maintenance and reduce future deficiencies resulting in improved long-term performance and lower life cycle cost
2	Warranties only incentivize the contractor to meet the bare minimum standards	Contractors are incentivized to achieve performance that exceeds expected performance

3 Examples of PBCs

PBCs have been used in a variety of industries like healthcare, rail maintenance and defense. Below we present examples of PBCs' usage in industry and summarize key takeaways.

3.1 Healthcare

An important recent development in the healthcare market in countries such as the US and the UK is the use of quality or outcome assurance in healthcare contracts. In such contracts, service providers (doctors, hospitals, health care clinics) are paid an incentive or extra compensation for improvements in medical outcomes. Although still at an early stage, the introduction of quality assessment and control in healthcare contracting is likely to have profound effects on the healthcare industry.

The need for PBCs in the healthcare industry stems from the problem of information asymmetry. Patients are unaware of the quantity and quality of treatment required, giving providers an informational advantage over the patients. Healthcare payers (insurance companies) mitigate the effect of this informational advantage of physicians over patients by contracting that payments will be based, in part, on measured output. For example, a substance abuse provider may receive an increase in funding from the state health department for reducing chemical dependancy of a greater than expected number of patients. Below, some examples of PBC usage in the healthcare industry are provided. All examples are adapted from [6].

Maine's Office of Substance Abuse
The state of Maine has used PBCs in basing funding of public substance abuse treatment providers on performance. Performance was measured based on the following three categories of metrics:

1. Efficiency - Measured by service delivery to primary clients.
2. Effectiveness - Measured by abstinence/drug free days prior to termination, reduction of primary drug use frequency, employement improvement, and reduction in the number of problems with employer/spouse.
3. Special Population Standards - Measured by the percentage of females among treated population.

The state of Maine also prescribed certain minimum performance standards that the providers were required to comply with. For example, efficiency standards required outpatient programs to deliver at least 90% of contracted units of treatment and at least 70% of those to primary clients. Introduction of PBC was correlated with measured performance and it was found that effectiveness (as defined by the state) improved.

Illinois's Quality Improvement Program
The Illinois Department of Public Aid introduced Illinois's Quality Improvement Program (QUIP) to improve nursing home resident quality by paying a premium over the basic reimbursement rate, for facilities that meet identified standards of care. Performance was measured in six areas -

1. Structure and environment

2. Resident participation and Choice
3. Community and family participation
4. Resident satisfaction
5. Care plan and,
6. Specialised intensive services

A general upward trend in percent of facilities meeting the identified standards was observed.

Quality care compensation system
US Healthcare (a health maintenance organization) developed a compensation model which increased payment to participating primary care physicians, hospitals and specialists depending on realized performance in providing improved quality of service. Based on this, a steady improvement of quality was observed. The performance metrics used were -

1. Quality review (measures of satisfaction from member surveys)
2. Comprehensive care package (membership size, schedules and available office hours, participating in US Healthcare sponsored continuing medical education)
3. Utilization (Number of hospital measured bed days, use of specialist care and emergency department services)

PBCs incentivize providers to refer patients to suitable providers, avoiding a bad treatment outcome and financial penalty. However, providers may also admit clients selectively and grant treatment only to those who are likely to improve. Providers may also try to game the system by reporting only the better treatment outcomes. Treatment outcomes that are subjective (e.g. mental health) pose challeges in implementation of PBCs.

3.2 Rail Maintenance

In-house contract between public infrastructure owner and production unit [7]
Swedish public rail infrastructure owner Banverkat negotiated a performance-based rail maintenance contract with its production unit to improve the service of its rail network. Performance was quantified by the number of technical errors and minutes of train delay, and were monitored by a third party. Incentives were provided for improvements over target values of the performance metrics which were weighted as 60% for delay improvement and 30% for error improvement. Outputs were measured before and after incentives were introduced and it was found that there was a reduction in delay by $10 \pm 5\%$ and the number of errors by $20 \pm 5\%$, at no additional cost to the owner.

It seemed at first that the effect of incentives reduced with time. Upon analysis of the data it was found that extreme weather conditions created a bias in the number of delays measured. To avoid an unfair penalty to the contractor, data collected during adverse weather conditions were not considered. Another bias found was that of incorrect registration of errors that disappeared without any contractor effort. Such errors were removed from the data and the contractor was not penalized for them.

Thus, in the rail maintenance industry PBCs have been implemented to improve quality of service. Two lessons that should be given attention in PBC design for the transportation industry are the following: (1) do not ignore external effects like adverse weather conditions, and (2) do not penalize contractors for errors that are not under their direct control.

3.3 Defense

Environmental PBC by the Air Force [8]
The Andrews Air Force base in Maryland was polluted due to contaminants that had penetrated deep into the groundwater. Traditional approaches to clean the area were attempted but did not work. The Air Force Center for Environmental Excellence (AFCEE) contracted BEM consulting under the first environmental PBC in the Air Force. BEM consulting tried a new type of advanced bioremediation technology on the groundwater, which worked miraculously well. Following this operation, the Air Force base was judged to be the most environmentally friendly base in 2002 and 2004. The initial project cost and time required for this cleanup were estimated at $1.5M and 26 years respectively. However due to the PBC and the contractor's innovative approach, the project was completed with $500K in expenditures in just 11 months.

Environmental PBC by the Army [9]
The Installation Restoration Program (IRP) was initiated by the US Army to complete clean up of its 1,080 installations by 2014. The Army employed PBCs in 2002 for its most technically challenging cleanups. The contractors were required to achieve specific objectives outlined in a performance-based statement and the Army no longer dictated how private firms perform cleanups, which allowed firms more flexibility and incentive to innovate. As a result, the projects were completed with a lower budget than planned leading to savings of about $130 million or approximately 30% of estimated costs.

3.4 Computer Maintenance

A major concern to users of computer systems is how to specify the desired level of reliability in maintenance contracts. Provisions that require a certain number of engineers to be present within a specified geographical vicinity or that the vendor respond to calls within a specified period of time place too much emphasis on the means by which performance may be affected, rather than the performance itself. For example, having two engineers in the neighborhood may not provide required level of performance if they are busy servicing other systems. In contrast, in a performance-based approach that tracks, for example, the percent up-time, the vendor is penalized if it fails to keep the system working for a pre-agreed amount of time. In this approach, the vendor has both the incentive and the flexibility to improve performance.

In [10], a framework for penalty computation is proposed based on the difference between the actual monthly down-time and the allowable monthly down-time for each hardware component. A disadvantage of this method is that a vendor may accept only limited liability for components supplied, even when these components are critical to system performance. For example, a defective processor can bring the whole system down but the vendor's liability is limited to the cost of the processor. This disadvantage is resolved by associating with each hardware component a parameter that reflects the importance or value of that component relative the performance of the entire system. For example, a central processor, which is critical to system operability, is given a value of 100% whereas a printer is given a value of 20%. An analogous approach can be used in PBCs for pavement construction, where weights can be assigned to different performance metrics based on the extent to which they affect pavement performance from road users' perspectives.

3.5 Highway Maintenance and Construction

Performance-Based Maintenance Contracting (PBMC) has been used for highway maintenance operations in the United States (Virginia, Texas, Florida), Canada, Australia and several South American countries. PBMC use is accelerating worldwide and overall it is found that PBMCs result in better outcomes at lower cost [11]. Below we highlight two examples from the US transportation construction industry.

In 1998, the District of Columbia Department of Public Works awarded a 5-year $69 million performance-based contract to a contractor for maintenance of 75 miles of the National Highway System (NHS) within the district [11]. Payments to the contractor included incentives and disincentives depending on achievement of performance standards. Over 170 performance metrics such as pavement markings, traffic signs, highway lighting, snow and ice control, etc. were covered in the contract and five composite measures were devised reflecting performance on assets and operations and response times to address issues. Each month an independent third party, along with District and contractor staff, inspected sections of the highway and rated each performance metric as poor, fair or good. Then a composite score was calculated. At the end of each year, an objective, comprehensive evaluation was conducted by the third party and the contractor was able to earn a variable award fee for excellent performance. The District observed significant improvement in the conditions of the contracted road assets. After the first year, performance rose from the high 20s to the low 80s (where performance was measured on a scale of 1 to 100).

In 1999, Texas Department of Transportation (TxDOT) entered into similar performance-based maintenance contracts, which involved sections of the Interstate highway with some of the heaviest traffic in the state [11]. Performance standards were developed for assets (pavements, bridges, roadsides) and operations (traffic operations, incident response, emergency repairs, etc.). Assets were rated for performance metrics like rutting, failures, litter, etc. on a scale of 1 to 5, where 5 is excellent and 1 is failed. Each metric was given a priority multiplier based on its relative importance and a composite score was calculated by adding the weighted scores. The projects were awarded based on lowest bid and monthly payments were calculated by multiplying the winning contractor's lump-sum bid by the monthly payment schedule percentage and making any deductions dictated by contractor performance. Here, service levels were found to decline initially and then started to rise. The contractor provided higher-quality ice and snow control than the agency previously did. It was also found that less inspection was required and the contractor was encouraged to be more innovative.

In the following year, TxDOT entered into four 2-year performance-based contracts, valued at around $7 million each, to upgrade and maintain the picnic and rest areas in the state. TxDOT established an evaluation process, a rating system, and an incentive/disincentive scheme to ensure that conditions improved and goals were met. Before the project the average rating score for the rest areas was 73% (rating scores ranged between 0% to 100%). TxDOT established the goal of increasing the average score across the state to 85%. For each day that a contractor scored above 92%, it received a 15% incentive payment of the normal daily pay. Contractors that scored 85% or lower received deductions in daily pay according to declining thresholds. After the first year of these contracts, TxDOT had paid incentives and assesed disincentives of nearly an identical amount of $\$246,000$. Average statewide rankings of facility conditions increased from 73% before the performance-based maintenance contracts to 91% at the end of the first year.

More recently, the construction aspects of highway projects have been incorporated into these performance-based contracts, extending PBMC to include construction value as well as maintenance. Examples from British Columbia and Michigan are described below.

British Columbia's Ministry of Transportation (MoT) decided to make improvements to the Sea-to-Sky highway, which is a 95-kilometre long section of highway between West Vancouver and Whistler, for the 2010 Winter Olympics. The purpose was to increase highway safety, reliability and capacity by highway widening and straightening and other measures. MoT decided to employ a 25-year performance-based public-private partnership contract. According to the terms of this contract, the contractor was expected to design and construct highway improvement on nearly two-thirds of the corridor, and then operate, maintain and rehabilitate the full corridor in keeping with performance standards of the contract. The following criteria were utilized by MoT to select the delivery model in the PBC -

1. Deliver the baseline improvements on time and on budget
2. Deliver additional highway improvements
3. Transfer appropriate risks to the private sector at appropriate prices;
4. Include incentives in the contract to achieve project performance objectives, maintain project schedule and budget, and address traffic management requirements
5. Achieve value for money

The PBC included performance incentive payments for adhering to traffic management standards (i.e., if the contractor exceeded the number and duration of stoppages or closures set out in the contract, payment would be reduced) and ensuring that the safety performance of the pavement exceeds expected safety performance. If the contractor failed to meet specified operational and maintenance standards and travel time delay experienced by road users, penalities would be levied. The allocation of risk was also divided based on which party would be more able to cost-effectively manage those risks with the contractor assuming most of the construction (time and cost overruns), financial, and traffic management risks and MoT taking on significant natural events (landslides) and regulatory risks. Value for money for this project was demonstrated through additional improvements provided in the PBC contract and the anticipated user benefits that flow from them. While the estimated cost of the PBC contract exceeded the expected cost if MoT had pursued a series of Design-Build contracts, various qualitative benefits were realized that demonstrate value for money. A more detailed analysis of this project can be found at [12].

In 2008, The Michigan Department of Transportation (MDOT) used a similar performance contracting for construction (PCfC) mechanism to select a contractor for construction on M-115, a rural highway [13]. Contractor bids were scored on six performance standards and on innovation, and the evaluation score was used to calculate a cost multiplier between 0.8 and 1, which would be multiplied by the contractor bid to represent its value. Contractors were also to provide a warranty bond to perform maintenance work when certain thresholds were met, defined in the request for bid. The contractor with the lowest adjusted bid was given the contract. The contractor won a total of $\$340,100$ in incentives by exceeding the performance standard requirements, and the high quality of work led to the project being labeled a success, with MDOT considering future applications of these contracting mechanisms.

3.6 Key Takeaways

Based on these examples, the following is a list of industry best practices to keep in mind when implementing PBCs for the transportation construction industry:

1. Clearly define performance metrics
2. Ensure that an outcome measurement infrastructure is in place
3. If possible, use 3rd party monitoring/assessment of performance
4. Permit leeway for uncontrollable risks (e.g., unanticipated temperature fluctuations)
5. Appropriately limit responsibilities and liability to reduce likelihood of disputes
6. Develop mutually agreed-upon benchmarks/targets at the beginning of the project
7. Ensure transparency in communications and data exchange

4 Performance-Based Specifications - Methodology

In this section we describe the PBS methodology for Hot Mix Asphalt (HMA) construction. All details have been adapted from [14], [15], [16] and [17]. The reason for devoting more attention to PBS contracts is that these contracts have several features in common with PBCs. Specifically, the calculation of expected performance may be done in the same way for both PBS and PBCs. However, two key differences should be noted. First, PBCs require an agreement on performance benchmarks and whereas they may be calculated using the prediction methods employed in PBS contracts, such an approach is not required for implementing PBCs. Second, actual performance is also predicted in PBS, but that is not the case in PBCs.

The general approach for distress prediction is based upon the MEPDG, but to perform computations quickly CFS have been developed. CFS predict a single value for each distress at the end of project design life. Monte-Carlo simulations are run on the CFS with the design specifications as input to predict the as-designed distress. The as-designed distress is then used to predict the remaining service life of the pavement. Similarly, Monte-Carlo simulations are run using as-built specifications as input to estimate the remaining service life. Predicted Life Difference (PLD) is calculated as the difference in service life predicted for as-designed and as-built specifications. Based on the PLD, either a penalty or a bonus (I/D) is assigned to the contractor on a lot-by-lot basis. These are weighted for different distress types and added to obtain total penalty/bonus. Finally a ride quality I/D based on the International Roughness Index (IRI) is added.

Figure 4.1 shows an example of the service life distribution of a pavement. The solid line represents the design specification, while the dotted lines are two examples of the predicted lives of actual construction. While the line to the left of the design specification clearly has a negative PLD, the PBS contract would define a rule by which the PLD would be defined for the other case. Figure 4.2 shows an example of an I/D scheme as a function of PLD. No incentives or disincentives are paid if PLD is close to 0, while larger deviations will have I/D tiers. Incentives are commonly capped, as in this example. In Section 4.1, we delve deeper into CFS for rutting and fatigue cracking distresses.

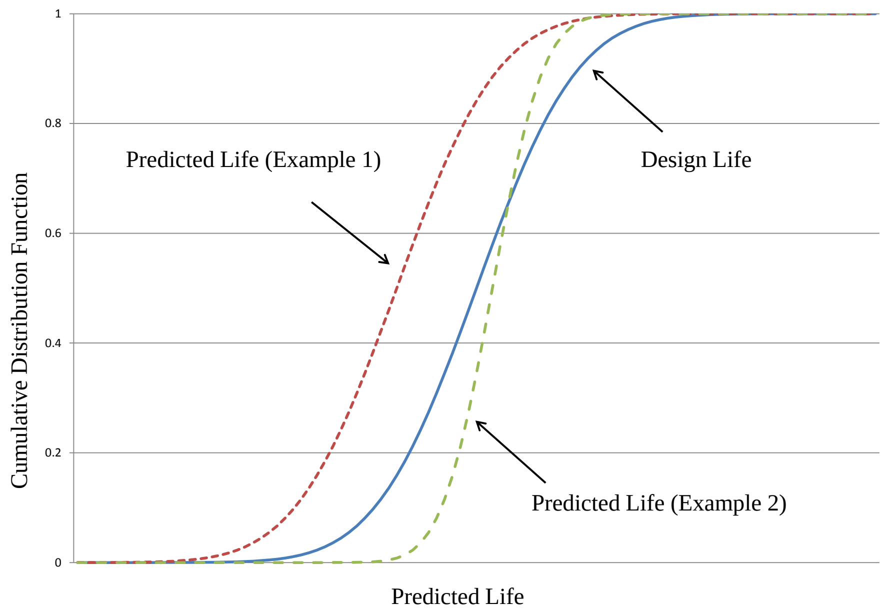

Figure 4.1: As-built and As-design Cumulative Frequency Distribution

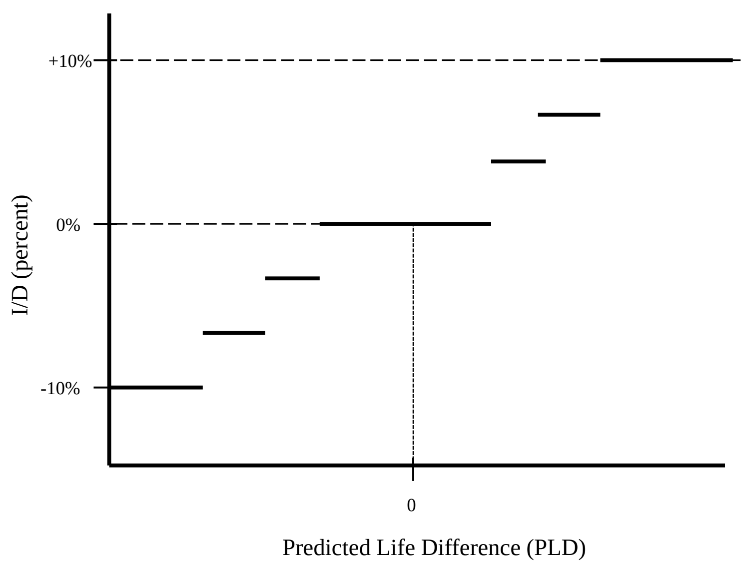

Figure 4.2: Example of an I/D Scheme

4.1 Performance Deterioration Models

At present due to the complexity of calculations, the MEPDG requires significant amount of time (months) for the Monte Carlo simulations. The major advantage of CFS methodology is that a probabilistic solution can be determined in a matter of minutes. As reported in the literature, CFS for prediction of rutting for asphalt pavements may be derived by the following steps -

1. Simulation runs of the MEPDG were done using various combinations of inputs and a database for each distress was created. An example of the matrix of inputs used in the MEPDG is shown is Table 4.1.

2. For a given structure, the dynamic modulus (E^*) of the HMA layer is obtained for each of the predefined climate and traffic conditions.

3. A relationship between the distress and dynamic modulus of the HMA is obtained, which can be used to predict the distress.

Table 4.1: Example of Distress Prediction Inputs (Source: [17])

Factor	**Sample Values**
Environmental sites	Hot: Phoenix, AZ, Cold: Grand Forks, ND
Design life (years)	20
Design traffic (ESALs)	2 x 10^6
HMA thickness (inches)	$1, 2, 3, 8, 20$
Vehicle speed (mph)	$0.5, 15, 45$
E^*	PG Grades (PG $82-10$,PG $64-22$), V_a (Air voids) and Vb_{eff} (Effective binder content) ($V_a = 4, Vb_{eff} = 8$)

This methodology is based on the following simplifying assumptions:

1. Traffic is represented by ESALs (Equivalent Single Axle Loads).

2. The methodology predicts one distress value at the end of the design life.

3. Design life of 20 years is used in these models.

4.1.1 Rutting

Rutting is observed when a longitudinal surface depression is observed in the wheel path. Extent of rutting is dependent on several factors like climatic conditions, asphalt mix properties and traffic conditions. Correspondingly, there exist closed form solutions for rutting as a function of effective temperature, effective dynamic modulus and traffic loading.

Witczak Predictive Equation Model
The dynamic modulus is a fundamental property of the HMA layer that is used in mechanistic design as calculated by the Witczak Predictive Equation [14] below:

$$\begin{aligned} \log(E^*) \quad = \quad & -1.249937 + 0.029232p_{200} - 0.001767(p_{200})^2 - 0.002841p_4 \\ & - 0.058097V_a - \frac{0.8022Vb_{eff}}{(V_a + Vb_{eff})} \\ & + \frac{(3.87197 - 0.0021p_4 + 0.003958p_{38} - 0.000017(p_{38})^2 + 0.00547p_{34})}{(1 + e^{-.603313 - .31335\log(f) - .395332log(\eta)})} \end{aligned} \tag{1}$$

E^* = Asphalt Mix dynamic modulus (in 10^5 psi)
η = Bitumen viscosity (in 10^6 poise)
f = Loading frequency (in Hz)
V_a = Air voids in the mix, by volume (%)
Vb_{eff} = Effective bitumen content, by volume (%)
p_{34} = Cumulative percent retained on the 3/4 inch sieve
p_{38} = Cumulative percent retained on the 3/8 inch sieve
p_4 = Cumulative percent retained on the No. 4 sieve
p_{200} = Percent passing through the No. 200 sieve

The frequency is calculated as a function of vehicle speed and the effective length (L_{EFF}) at the mid depth of an HMA layer [14],

$$f = 17.6\frac{Speed}{L_{EFF}} \tag{2}$$

E^* is thus a function of Asphalt Mix properties like binder stiffness and voids filled with bitumen (VFB). Binder stiffness in turn depends on effective temperature (climatic conditions).

Climactic Conditions
The effective temperature (T_{eff}) is defined as a single test temperature at which an amount of distress would be equivalent to that which occurs from the seasonal temperature fluctuation throughout the annual temperature cycle. The effective temperature concept negates the necessity to conduct cumulative incremental damage throughout a change in annual temperature conditions.

For rutting [14] propose that,

$$\begin{aligned} T_{eff} \quad = \quad & 14.62 - 3.361\log(f) - 10.940z + 1.121(MAAT) + 1.718(s_{MMAT}) \\ & - 0.431W + 0.33S + 0.08R \end{aligned}$$

and for fatigue cracking,

$$\begin{aligned} T_{eff} &= -13.995 - 2.332Ln(f)^(.5) + 1.006(MAAT) + 0.876(s_M MAT) \\ &\quad - 1.186W + 0.549S + 0.071R \end{aligned}$$

In the above expressions, the following abbreviations are used:
T_{eff} = Modified Witczak effective temperature (deg F)
z = Critical depth (inch)
f = Loading frequency (in Hz)
$MAAT$ = Mean annual air temperature (deg F)
s_{MMAT} = Standard deviation of the mean monthly air temperature
R = Annual cumulative rainfall depth (inches)
S = Mean annual percentage sunshine (%)
W = Mean annual wind speed (mph)

Predicted Rutting Expression
Using the effective temperature and frequency values corresponding to the project, E^* values were calculated and related to the predicted MEPDG rutting. A highly correlated power function was found to very accurately define the relationship between E^* and predicted MEPDG rutting. For particular climate and traffic conditions,

$$R_{D_i} = a_i E_i^{*^{b_i}} \tag{3}$$

R_{D_i}= Predicted rut depth in each sublayer i (inch)
E_i^*= Effective Dynamic Modulus for each sublayer i
a_i, b_i= Regression coefficients set for each sublayer i

The total Rut depth is obtained by summing the predicted rut depth in each sublayer of the HMA layer as,

$$RD = \sum_{i=1}^{n} \epsilon_{p_i} \Delta h_i \tag{4}$$

RD = Total HMA layer rut depth
ϵ_{p_i} = Plastic strain
Δh_i = Thickness of HMA sublayer i
n = Total number of sublayers

Recent research shows that for most pavements, the base and subgrade properties can affect the performance of the pavement. In [18], a modification to the rutting model is proposed where total rutting is expressed as a function of base and subgrade rutting. We report the modified equation below.

$$\text{Total Rutting} = \text{Rutting}_{AC} + \text{Rutting}^*_{\text{base}} + \text{Rutting}^*_{\text{subgrade}}, \tag{5}$$

Rutting_{AC} is the rutting in the HMA layer and $\text{Rutting}^*_{\text{base}}$, $\text{Rutting}^*_{\text{subgrade}}$ are the original MEPDG predictions for these layers from which the predicted rutting in the first month is subtracted.

The above expression clearly indicates that the base and subgrade rutting also affect total measured rutting. This implies that when predicting performance for benchmarking purposes, it is important to use the correct approach. However, once a benchmark performance is established, our approach does not depend on the functional form of rutting expression. In PBCs, the contract is based on the observed pavement distress. The contractor is responsible for choosing an appropriate mix and construction methods. Given an optimally designed I/D scheme, the contractor will take steps to minimize observed rutting, which by default includes rutting in all layers.

Prediction Accuracy

Figure 4.3 shows that the rutting prediction is excellent and very relevant to the MEPDG prediction. The regression coefficient (R^2) is 0.996, which indicates an excellent correlation between the two prediction methods.

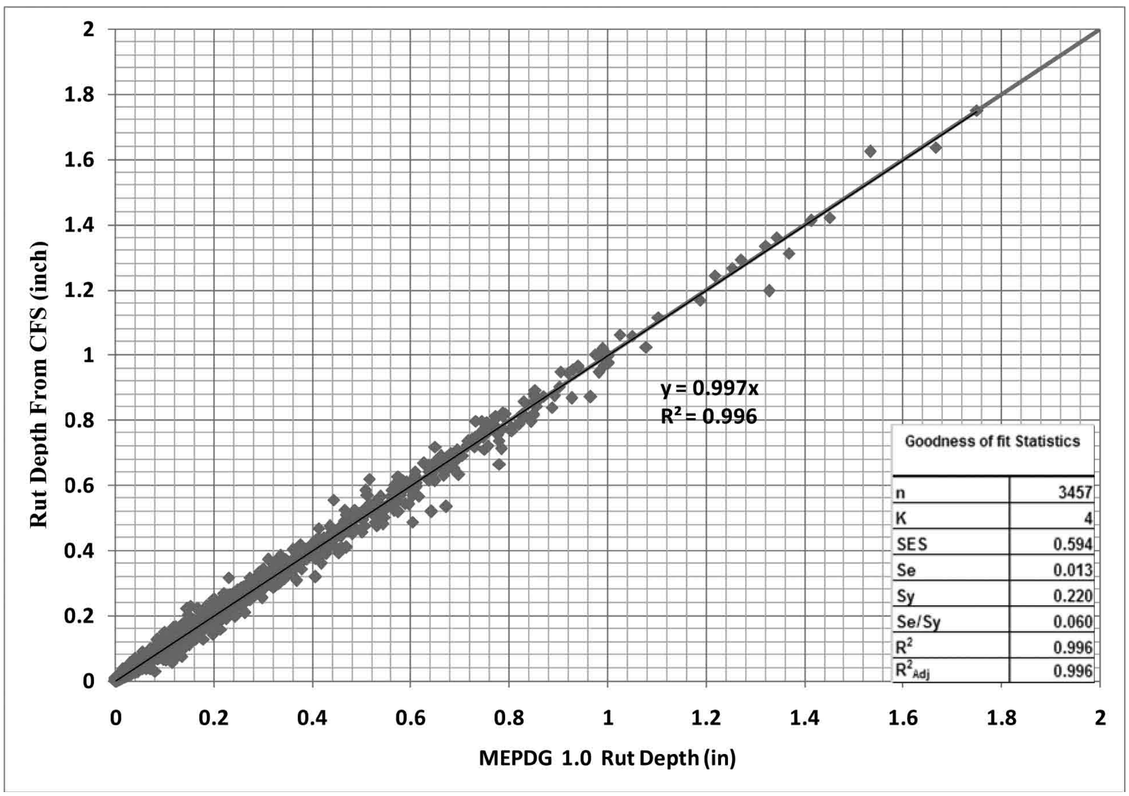

Figure 4.3: Prediction Accuracy of Rutting CFS (Source: [14])

4.1.2 Alligator/Fatigue Cracking

Similar to rutting, a general comprehensive model to predict fatigue damage was also developed. The CFS of fatigue cracking is expressed as,

$$FC = \frac{100}{1 + e^{C_1 + C_1^* + C_2 + C_2^* \log(D)}}, \tag{6}$$

where FC = Fatigue cracking (% of lane area)

C_1 = C_2 = 1.0
C_1^* = $-2C_2^*$ = $-2 * (-2.40784 - 39.748(1 + h_{ac})^{-2.856})$
h_{ac} = Thickness of asphalt layer (inches)

$$D = \sum_{i=1}^{T} \frac{n_i}{Nf_i} \tag{7}$$

D = Fatigue damage
T = Total no. of computational periods
n_i = Actual traffic for period i
Nf_i = Allowable failure repetitions under conditions prevailing in period i

Here total fatigue damage is the cumulative damage over T computational periods. The allowable number of repetitions to failure (N_f) in any period is calculated as follows [17]:

For $h_{ac} \leq 3$ inches (*thin model*),

$$\begin{aligned}
\log(N_f) \quad = \quad & 8.3014 - \{[(-0.0096 \log(h_{ac})^2 - 0.0756 h_{ac} + 0.0438) \log(E^*) \\
& - 0.5414 \log(h_{ac}^2) + 1.4319 \log(h_{ac}) - 1.0252] \log(E_{cf}^2) \\
& + [-0.0208 \log((E^*)^2) + 0.7040 \log(E^*) - 4.1771 \ \log(E_{cf}) \\
& + [-4.1659 \log(h_{ac}^2) - 3.0733 \log(h_{ac}) - 6.4418] \log(VFB)^2 \\
& - [1.5883 \log(h_{ac}^2) - 2.8014 \log(h_{ac}) - 9.2885] \log(VFB) \\
& - 0.1177 \log((E^*)^2) + [0.0681 h_{ac}^2 - 0.3789 h_{ac} + 0.8989] \log(E^*) + 2.9330\}
\end{aligned}$$

For $h_{ac} \geq 3$ inches (*thick model*),

$$\begin{aligned}
\log(N_f) \quad = \quad & 8.3014 - \{[(-0.0645 \log(h_{ac})^2 - 0.0144 h_{ac} + 0.0416) \log(E^*) \\
& - 0.6003 \log(h_{ac}^2) + .7046 \log(h_{ac}) - 1.0276] \log(E_{cf}^2) \\
& + [-0.0218 \log((E^*)^2) + 0.6280 \log(E^*) - 3.2499 \ \log(E_{cf}) \\
& + [-28.9186 \log(h_{ac}^2) - 51.9588 \log(h_{ac}) + 12.7671] \log(VFB)^2 \\
& - [15.8844 \log(h_{ac}^2) - 28.6128 \log(h_{ac}) - .9160] \log(VFB) \\
& - 0.1792 \log((E^*)^2) + [0.0024 h_{ac}^2 - 0.1009 h_{ac} + 1.2623] \log(E^*) + 1.4613\}
\end{aligned}$$

h_{ac} = Thickness of asphalt layer (inches)
E^* = Effective dynamic modulus
E_{cf} = Composite foundation modulus (ksi)
VFB = Voids filled with bitumen

Once N_f is determined, the fatigue damage at the ESALs of interest can be calculated using Equation 7. In addition fatigue cracking can be calculated from predicted damage by using the MEPDG transfer function shown in Equation 6. Note that the parameters E^* (Effective dynamic modulus) and h_{ac} (thickness of HMA layer) are common to both rutting and fatigue cracking distress prediction models.

Prediction Accuracy
Figure 4.4 shows very good agreement between the percentage fatigue damage predicted with the developed methodology and MEPDG. The adjusted R^2 is 0.982.

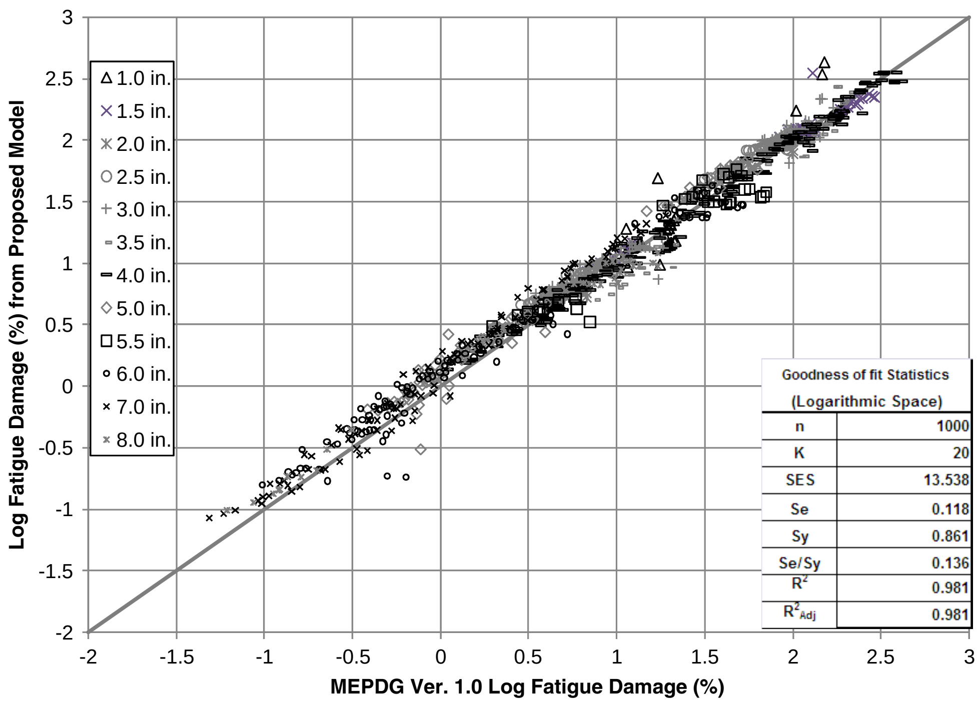

Figure 4.4: Prediction Accuracy of Fatigue Cracking CFS (Source: [17])

Recent research has also focussed on improving the CFS expressions to make predictions more accurate. Significant progress is also being made on the low temperature cracking prediction model using finite element methods to determine the thermal cracking (TC) rate in the pavement [19]. Tests other than dynamic modulus are also being employed to ensure acceptable rutting resistance of HMA paving mixtures.

The above-mentioned methods can help improve prediction and inform contractors how to affect controllable properties of the mix to realize desired performance. From the contracting viewpoint, these methods affect the setting of the benchmark performance against which realized performance is compared. The key problem we are attempting to address in the current project is that of mechanism design – i.e. the determination of the size of incentive/disincentive payments as a function of the discrepancy between expected and realized performance, and contractor response (in terms of bid parameters). The owner and the contractor need to agree that performance prediction models are reasonably accurate. However, the particular method used to predict performance does not affect the mechanism design problem because the payments depend on realized performance and benchmark performance (which may be based on prediction models) is an input to such models.

4.2 Relationship Between Contractor Effort and Performance

In Section 4.1, we gave an overview of the rutting and fatigue cracking performance deterioration models. Both rutting and fatigue cracking can be represented by closed form solutions which are nearly as accurate as MEPDG predictions. These may be used to establish benchmarks. Below we give an example to show that the contractor can indeed affect performance of the pavement.

4.2.1 An Example

Two factors that influence Thermal Cracking (TC) are binder type and volumetric properties (asphalt content and air voids). In Figure 4.5, we illustrate the development of TC over time for two different mixes in an example from [16]. Figure 4.5 shows that asphalt pavements with higher effective binder content Vb_{eff} can endure longer without experiencing any significant TC fractures. Thus, contractors who control the type and quantity of binder to use in the as-built mix can affect the extent of observed distress.

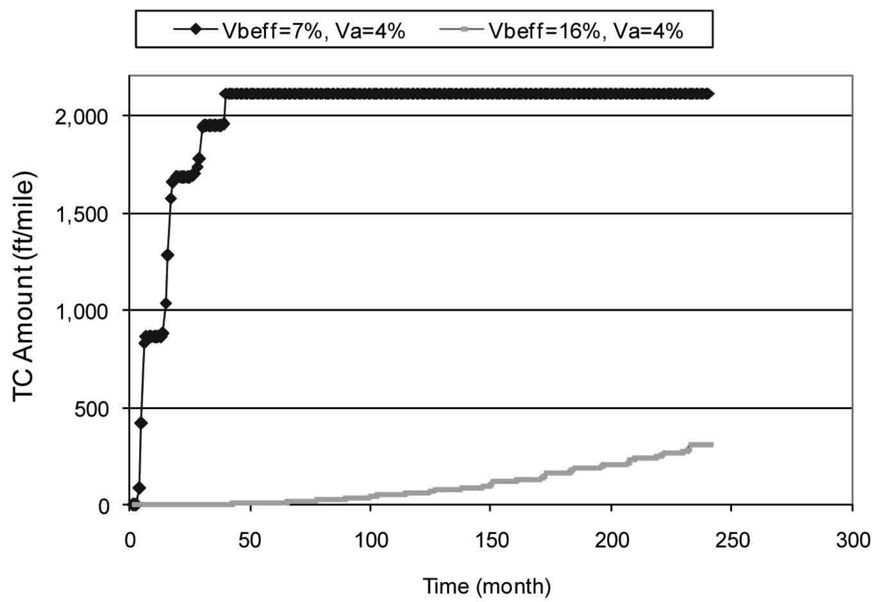

Figure 4.5: Contractor Effort Reduces Thermal Cracking (Source: [16])

4.2.2 Contractor Controllable Factors

We list the factors that contractors can control to affect the pavement performance in Table 4.2.

Table 4.2: Asphalt Mix Properties Under Contractor Control

1	V_a = Air voids in the mix, by volume (%)
2	Vb_{eff} = Effective bitumen content, by volume (%)
3	p_{34} = Cumulative percent retained on the 3/4 inch sieve
4	p_{38} = Cumulative percent retained on the 3/8 inch sieve
5	p_4 = Cumulative percent retained on the No. 4 sieve
6	p_{200} = Percent passing the No. 200 sieve
7	η = Bitumen viscosity (in 10^6 poise)
8	Amount of Recycled Asphalt Pavement (fractionated RAP)
9	Shingle byproduct content, by weight
10	Percent crushed particles
11	Mix moisture content
12	Aggregate specific gravity (i.e. taconite, steel slag)
13	Concentration of Anti-strip additives

By changing the factors listed above the contractor can control the asphalt mix properties through the HMA dynamic modulus (E^*), which in turn helps reduce the predicted rut depth. In addition, the contractor can also control the thickness of the pavement as well as the timing and extent of interventions. Figure 4.6 shows a plot between the dynamic modulus and predicted rut depth.

We note that for a unit increase in the dynamic modulus, the extent of reduction in predicted rutting is lower at higher dynamic moduli. Thus, the contractor sees diminishing returns when increasing the dynamic modulus.

The contractor can also perform a variety of maintenance activities over the course of the pavement's life. These can be classified as either preventive maintenance, which is performed to reduce future failures and increase the functional life of the pavement, or as corrective maintenance, which is performed after a deficiency occurs in the pavement. In Table 4.3, we highlight different types of maintenance actions along with the treatment category. Actions in the preventive maintenance category are performed when pavements do not have severe structural deficiencies and require relatively smaller effort and cost than rehabilitation and reconstruction treatments, which are done when the pavement is near failure. By designing PBCs in the right way we can ensure that the contractor is incentivized to perform preventive maintenance before the deficiencies snowball into damage requiring costly corrective actions.

Depending on the type and extent of intervention the contractor can reduce distress. Here, too, it would be reasonable to assume that the contractor will observe diminishing returns for the amount of intervention effort applied, at least within a treatment category.

4.2.3 Variability in Construction Practices

A contractor may reduce variability in construction practices and material properties to affect long-term pavement performance. In Figure 4.7, we see how reduction in variability of Recycled Asphalt (RAP) content increases reliability.

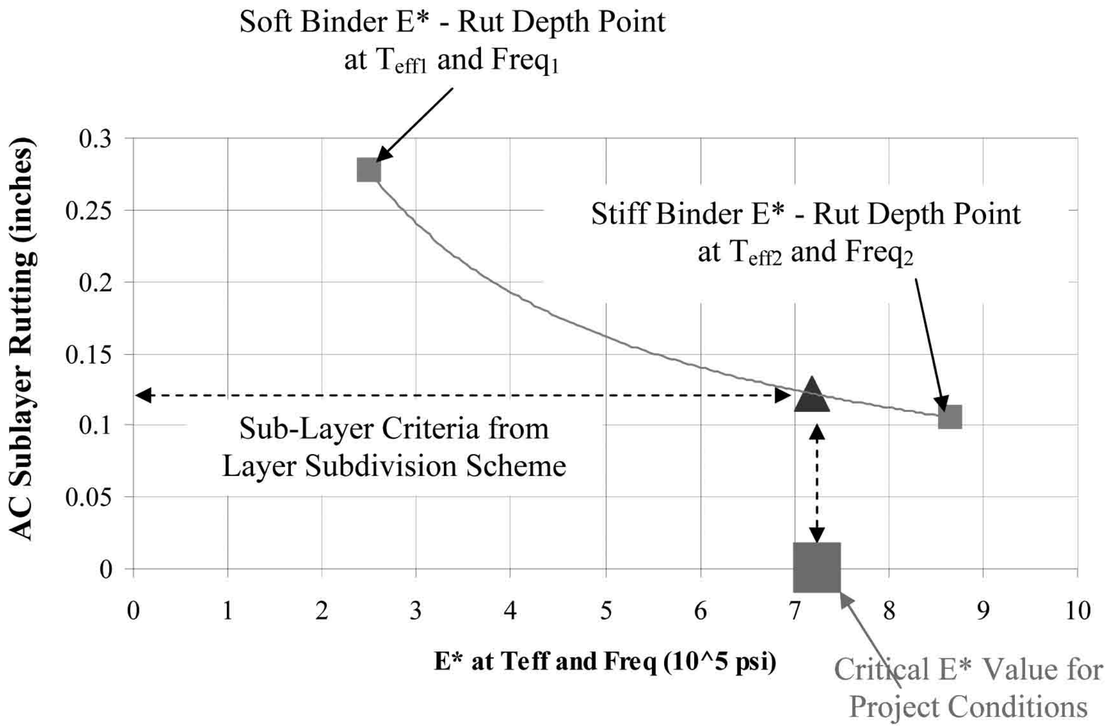

Figure 4.6: Contractor Effort Reduces Rutting (Source: [14])

Lowering variability in cement content and time between mixing and compaction also ensures high quality work. Compaction should be performed as soon as possible after mixing to achieve maximum dry density on the field [21].

Intelligent compaction equipment measures and records the quality of compaction during the compaction process. The compactors force changes in real time to increase compaction where needed, while preventing over compaction. Thus, use of this type of equipment and adherence to rolling patterns by the contractor increases compaction uniformity and reduces variability in built specifications. Another area where the contractor can reduce variability is the temperature at which the mat is placed and rolled. Final rolling should be performed when the mat has cooled to the degree that few or no roller marks are left by the roller and optimal densification can be accomplished. The contractor can also use tools to create a tapered edge like the Michigan wedge joint that results in higher density and longer lasting longitudinal joints, as well as statistical process control techniques such as Percent Within Limits to further reduce variability.

Table 4.3: Possible Maintenance Actions (Source: [20])

Treatment	Treatment Category	Effect
Do nothing	No action	None
Crack Treatments		
Clean & Seal	Preventive Maintenance	Correct minor deficiencies
Saw & Seal	Preventive Maintenance	Correct minor deficiencies
Rout & Seal	Preventive Maintenance	Correct minor deficiencies
Crack Filling	Preventive Maintenance	Correct minor deficiencies
Surface Treatment		
Fog Seal	Preventive Maintenance	Restore surface, extend life
Seal Coat	Preventive Maintenance	Restore surface, extend life
Double Chip Seal	Preventive Maintenance	Restore surface, extend life
Slurry Seal	Preventive Maintenance	Restore surface, extend life
Micro surfacing	Preventive Maintenance	Restore surface, extend life
Thin Hot-Mix Overlay	Preventive Maintenance	Extend life
Functional Overlays and Reclamations	Rehabilitation	Major surface upgrading or replacement
Hot Mix Asphalt Structural Overlay	Rehabilitation	Major surface upgrading or replacement
Asphalt Pavement Replacement	Reconstruction	Construct new pavement

4.3 Inventory of Performance Metrics

In Table 4.4, we list suggested performance metrics that can be used to evaluate contractor performance by STAs.

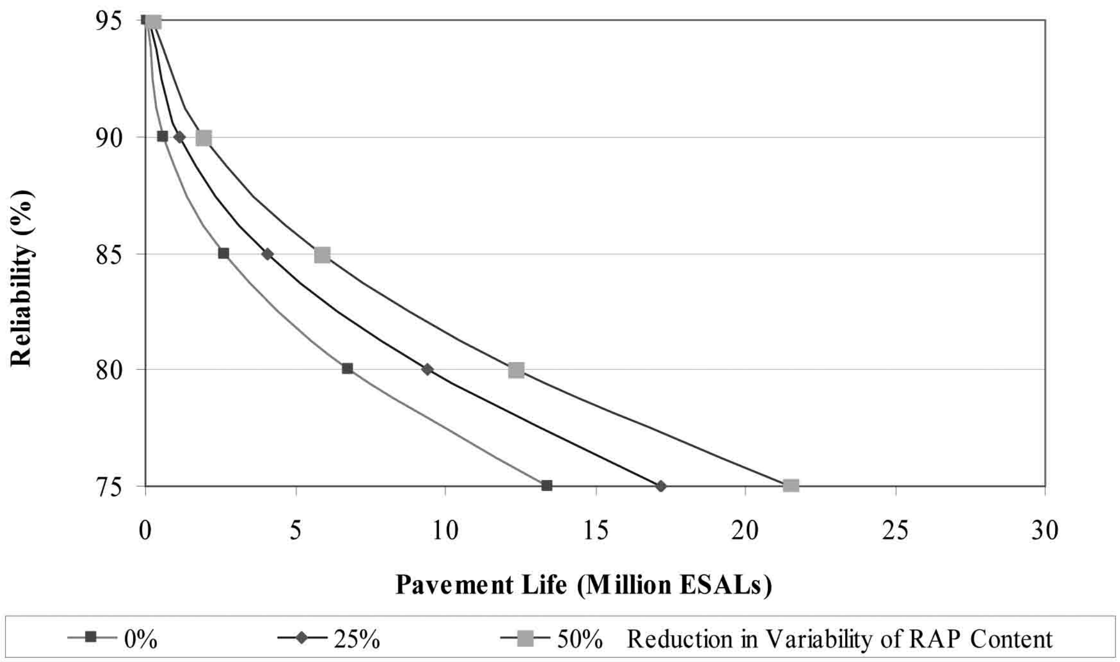

Figure 4.7: Less Variable RAP Content Improves Pavement Life (Source: [21])

Table 4.4: Performance Metrics (Source: [22])

S No.	Performance Metric	Description
1	Transverse Cracking	Cracks that are predominantly perpendicular to the pavement centerline
2	Longitudinal cracking	Cracks that are predominantly parallel to the pavement centerline
3	Multiple/Block Cracking	Pattern of cracks dividing pavement into approximately rectangular blocks
4	Alligator/Fatigue cracking	A series of interconnected cracks forming many-sided, sharp-angled pieces, six inches or less in size
5	Rutting	A longitudinal surface depression located in the wheel path
6	Roughness	Total anticipated vertical movement of vehicle accumulated over the length of the section

5 Design of Performance-Based Contracts

This chapter focuses on quantifying the effect of an STA's choice of incentive and disincentive (I/D) scheme on contractors' construction and maintenance efforts in performance-based contracts (PBCs). In Section 4 we examined a model for evaluation of performance and I/D payments, introduced in the context of PBS contracts. Now we turn to the question of how to design PBCs effectively.

We focus on how a profit-maximizing contractor who wins the bid might choose its maintenance effort at each decision epoch. Maintenance effort refers to the choice of maintenance activities that the contractor decides to perform at each decision epoch (see Table 4.3). Decision epochs consist of the points in time at which pavement performance is measured, I/D payments are made, and the contractor may take action if warranted by the pavement condition. For applications of the PBC method in the Minnesota Department of Transportation (MnDOT) context, researchers expect that decision epochs will arise once every year. We represent the contractor's problem in mathematical models that select maintenance activities at each decision epoch to maximize the contractor's expected profit-to-go over the remainder of the contract. Different activities entail different costs and lead to different levels of improvement in pavement performance.

For ease of modeling, we assume that the pavement performance belongs to one of a finite number of discrete states. In particular, the states are indexed by $i \in \{0, \cdots, s\}$ where zero represents the worst state and s represents the best. The value of s may be arbitrarily large to accommodate a variety of different performance measurement schemes. For example, if performance of the pavement is judged based on the Surface Rating (SR), a weighted index between 0.0-4.0, then the discrete states used in our model can be interpreted as discrete divisions of the SR index. A valid division may map state indexed zero to a rating between 0.0-0.5, state indexed one to 0.5-1.0, and so on until a rating between 3.5-4.0 denotes the best state (in this case $s = 7$).

In the remainder of this section, we characterize a profit-maximizing contractor's optimal sequence of decisions in response to the I/D scheme selected by the agency for each realized pavement state and time epoch. The collection of such decisions is referred to as an optimal policy. Thereafter, we propose two I/D schemes that lead to different policies. Finally, we comment on the impact of I/D payments on contractors' bids.

5.1 Notation and Assumptions

Let t denote an arbitrary decision epoch with pavement state i. Then the sequence of events is as follows. The contractor receives an I/D payment at time t, which depends on t and i. Also, at this point in time, the contractor decides how much, if any, effort to exert to improve the state of the pavement. Choice of effort is tantamount to the selection of a target state $j \geq i$. The cost of this effort increases with the amount of improvement desired. For example, if a small improvement is desired, then a cheaper Rout and Seal type of maintenance effort may suffice. However, if a significant improvement in pavement state is desired, then the contractor may need to undertake a costly effort like a functional overlay. Clearly, the amount that a rational contractor should be willing to spend on improving the pavement depends on the value of I/D payments it expects to receive over the remainder of its contract with the agency.

Given initial state i, target state j and decision-epoch t, we model the contractor's cost by the function $c_t(i, j)$, where $j \geq i$. It is straightforward to see that $c_t(i, i) = 0$. This represents the case when the contractor decides to do nothing in period t. Consistent with construction management literature, we assume that $c_t(i, j)$ is increasing in j for each fixed i. This follows from the observation that a higher performing target state requires greater effort. Similarly, for each fixed target state j, the effort cost is higher when i is smaller. That is, it is more expensive to go to the same target state from lower performing initial states. The cost of effort is dependent on the initial state, the target state, and the decision-epoch index.

After the contractor chooses its effort and takes pavement maintenance action, the target pavement state reached is denoted by j. Between t and the next decision epoch $(t + 1)$, pavement state may deteriorate due to environmental fluctuations and normal traffic loading. Recall that at $(t+1)$, pavement state is measured again, I/D payments are made, and the contractor has an opportunity to choose its effort, which will affect pavement state in all subsequent periods. For each time index t, pavement deterioration is represented by a matrix of transition probabilities P^t_{jk}s, where j is the target state reached in period t and $k \in \{0, 1, \cdots, j\}$ is the state realized at the start of period $(t + 1)$ on account of deterioration. We assume that the probability of realizing each state k is known. Note that j is greater than or equal to i, k is less that or equal to j, and i, j, k are integers between 0 and s.

The agency chooses a schedule of incentive and disincentive payments for each i and t. These amounts are specified before contractors bid on the project. We denote the I/D payments by $h_t(i)$. Note that the amount of incentive/disincentive that a contractor receives for realizing a particular state in a particular period depends not only on the state, but also on the time index. Specifically, we recognize that the agency may give higher incentives in later periods to encourage the contractor to exert more effort after the pavement starts to age.

Our models allow a great deal of flexibility in the agency's choice of $h_t(i)$. Two such examples are presented next. Either approach can lead to similar results in terms of contractor response. That is, the final choice of an I/D function, subject to certain requirements, can be based on agency preferences. In the first example, a hypothetical agency first identifies a minimum acceptable performance level in each period as $\bar{i_t}$. It pays the contractor according to the amount by which realized pavement state differs from $\bar{i_t}$. If the realized performance i is less than the acceptable performance, $h_t(i)$ may be negative, and if the the realized performance is greater than the acceptable performance, $h_t(i)$ may be positive. Negative values result in a penalty to the contractor and positive values imply incentive payments. The size of both the penalty and the incentive may depend on the magnitude of difference between i and $\bar{i_t}$, and the decision-epoch t. In this scheme, $h_t(i) = 0$ if the realized performance is equal to the acceptable performance. In the second example, a hypothetical agency bases its payments only on realized state. That is, no minimum performance level announced for each decision epoch. If the agency were interested in linear payment schemes, then these two examples are captured by a payment function of the type $h_t(i) = h_t \cdot (i - \bar{i_t})$. The first example arises when $h_t = h$ and and all values of $\bar{i_t} = 0$ are not zero, and the second case arises when $\bar{i_t} = 0$ for all t. However, in this case, h_t varied by t.

In the models presented in this report, we assume that the probability that the pavement state will deteriorate from state j to state k between two consecutive observation intervals, say t and $(t + 1)$, is independent of the time index of the observation. We refer to these probabilities as the state transition probabilities and represent them by P_{jk}. Put differently, we assume that $P^t_{jk} = P_{jk}$ for all t. Also, for the sake of concreteness, we model the extent of deterioration in each period via a process of random shocks A, which can take any non-negative integer value. For example, if the target state in a period is j and the size of the shock is 0, then there is no deterioration and

$k = j$; if the size of the shock is 1, the pavement deteriorates to state $j - 1$; and so on until shock equals or exceed j, at which point the resulting state is 0. Mathematically, this means that $P_{jk} = P(A = j - k)$, for $1 \leq k \leq j$, and $P_{j0} = P(A \geq j)$. Suppose $a_m = P(A = m)$ is used to denote the probability that the size of the shock is m. Then, this implies

$$P_{jk} = \begin{cases} a_{j-k} & \text{if } 1 \leq k \leq j \text{, and} \\ \sum_{\ell=j}^{\infty} a_\ell & \text{if } k = 0 \text{, and} \\ 0 & \text{otherwise} \end{cases}$$

Models with non-stationary transition probabilities and more complicated deterioration processes than the one described in this report are topics for future research.

5.2 Model Formulation & Assumptions

We use $\Pi_t(i)$ to denote the contractor's expected profit from period t onwards if its initial state is i, and $j_t^*(i)$ to denote contractor's decision, i.e., the optimum target state for each (i, t). Parameter α denotes the discount factor. Next, we formulate and solve the contractor's problem of finding an optimum target state $j_t^*(i)$ for each (i, t). Our goal is to identify an optimal policy, i.e. a set of decision rules that allow the contractor to determine its optimal action at each t, given initial state i. The contractor's profit-to-go function for each $t \in \{1, 2, ...n - 1\}$ and $i \in \{0, 1, ..., s\}$ is written as follows.

$$\Pi_t(i) = h_t(i) + \max_{j \geq i} g_t(i, j), \tag{8}$$

where

$$g_t(i, j) = \Big(- c_t(i, j) + \alpha \sum_{0 \leq k \leq j} P_{jk} \Pi_{t+1}(k) \Big) \tag{9}$$

is defined for every $j \geq i$. The first term in (8) is the incentive/disincentive that the contractor is paid in period t and the second term is the maximum future profit from all possible states j that the contractor may target. In equation (8), $j_t^*(i) = \arg\max_{j \geq i} g_t(i, j)$ is calculated for each i and t during the process of evaluation of contractor's optimal profit function. In the final period, n, the contractor simply receives an incentive/disincentive and makes no effort i.e.,

$$\Pi_n(i) = h_n(i). \tag{10}$$

In the remainder of this report, we focus on characterizing the contractor's optimal strategy without performing any numerical analysis. In order to derive such results, we make certain assumptions about the choice of the I/D scheme and cost functions. These assumptions are consistent with construction management literature, and therefore represent plausible scenarios. They may not hold in some situations. In all those cases, our model formulation can be solved numerically for given cost and I/D functions, to derive the contractor's optimal strategy for each state i and each time epoch t.

The following is a list of assumptions about cost and reward functions that we believe would apply to the vast majority of problems faced by agencies who wish to use PBC paradigm. For agencies considering the use of PBCs, it would be relatively straightforward to use historical costs data to verify if the assumptions listed below are reasonable. Agencies may also benefit from obtaining historical cost data from contractors.

1. The difference $c_t(i, j+1) - c_t(i, j)$ is non-increasing in j for all $j \geq i$ and each fixed i. This means that although cost increases as the contractor tries to achieve higher states, the rate of increase of cost is non-increasing. Put differently, for some $j > i$ and initial state i, the incremental cost of going from state j to state $(j+1)$ is no more than the incremental cost of going from state $(j-1)$ to state j. A graphical representation of this cost function is shown in Figure 5.1. In this figure, $j \in \{i, i+1, \cdots, s\}$. This type of cost function arises out of economies of scope as explained by the following scenario. Suppose maintenance effort entails some fixed cost of mobilizing crew and making arrangements for traffic management, which are not affected significantly by the type of maintenance activity. In that case, whereas costs may increase when greater effort is exerted, they may not increase at an increasing rate.

2. Cost $c_t(i, s)$ is either monotone increasing or monotone decreasing in t for each fixed i. It is necessary to make this assumption because no analytical results can be obtained if costs are not monotone. For example, if maintenance costs are expected to rise and fall across periods, then it would not be possible to say anything definite about the contractor's optimal strategy. In most practical settings, costs generally rise with time. Therefore, we expect this assumption to hold in many cases.

3. Incentives are greater when pavement is in better condition, i.e. $h_t(i+1) \geq h_t(i)$. This implies that the amount of incentive/disincentive that the contractor receives increases as the initial state i of that period increases. I/D schemes of this sort encourage contractors to target higher states which represent better performance.

4. The difference $h_t(i+1) - h_t(i)$ is non-decreasing in i. This means that the incremental benefit to a contractor of achieving state $i+1$ relative to state i is at least as much as the incremental benefit of achieving state i relative to state $i-1$. Non-decreasing incentives serve to encourage contractors to target higher performing pavement states. Agencies should be willing to offer such incentives because greater contractor investment toward pavement preservation activities is the reason for undertaking PBCs in the first place. A graphical representation of this I/D scheme is shown in Figure 5.2.

5.3 Results

In this section, we present a characterization of the contractor's optimal policy. Our goal is to show that there exist I/D schemes that can incentivize the contractor to achieve target pavement states desired by the agency. The results are presented in two theorems and related corollaries. For clarity of exposition, we introduce notation i_t^* to denote a threshold in period t such that for all initial states below the threshold, the contractor has a common optimal action and for all initial states at or above the threshold, the contractor has another common (possibly different) optimal action. Formally, $i_t^* = \{i \in 0, 1, ..., s$ such that $g_t(i-1, s) - g_t(i-1, i-1) < 0, g_t(i, s) - g_t(i, i) \geq 0\}$, or vice versa. In this section, we use the term increasing (resp. decreasing) interchangeably with non-decreasing (resp. non-increasing). All proofs are presented in the Appendix.

Before presenting our main result, we define quantities $\Delta f(i) \doteq f(i+1) - f(i)$ and $\Delta^2 f(i) \doteq \Delta f(i+1) - \Delta f(i)$ and a series of inequalities. These inequalities are requirements placed upon the reward function to achieve certain contractor response.

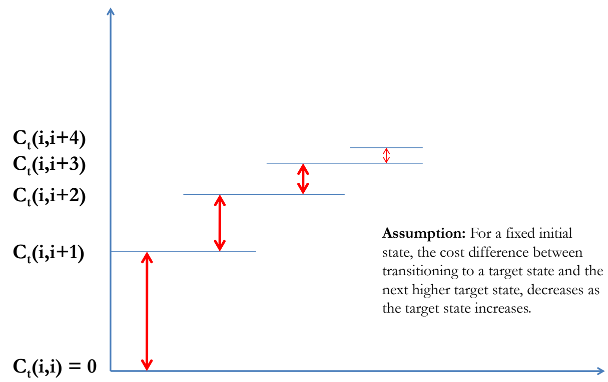

Figure 5.1: A Cost Function

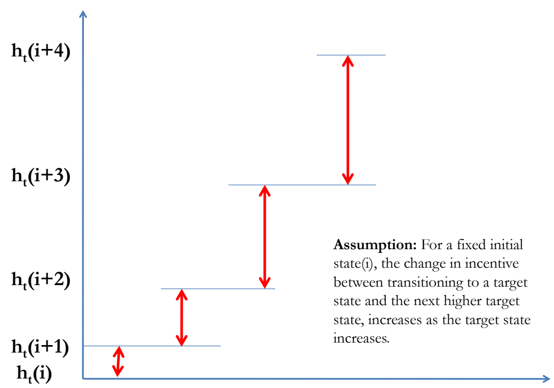

Figure 5.2: An I/D Function

I1: -$\Delta c_t(i,s) \geq \alpha \sum_{0\leq k\leq i} a_{i-k}\{h_n(k+1)-h_n(k)\} \quad t=1,...,n-1$

I2: $\Delta^2 h_t(i) - \Delta^2 c_t(i,s) \geq 0 \quad t=1,...,n-1$

I3: $\Delta h_{t+1}(i) - \Delta h_t(i) \geq \max\{0, \Delta c_{t+1}(i,s) - \Delta c_t(i,s)\} \quad t=1,...,n-2$

I4: $\Delta h_n(i) - \Delta h_{n-1}(i) \geq -\Delta c_{n-1}(i,s)$, $(1-\alpha)\Delta h_n(i) \geq \Delta h_{n-1}(i)$

Inequality I1 means that if it is optimal for the contractor to target state s from state $(i+1)$, then it is also optimal to do so from state i. I2 ensures that the rate of increase of the incremental differences in reward are greater than or equal to the rate of decrease of incremental differences in the cost when the target state is s. I3 requires that the rate of increase in incentive payment should increase with time. Specifically, if the rate of decrease in cost for targeting state s is increasing in time, then the incentive payment rate must increase faster than this.

Theorem 1 *Optimal contractor action in every period is either to stay at the initial state i or bring the target state to s. In particular, if the cost of targeting state s from each fixed initial state is* *decreasing in time and the incentive scheme satisfies inequalities I1–I4, then there exist thresholds i^*_t such that if initial state is below the threshold, the contractor stays at the initial state and if the initial state is at or above the threshold, then it is optimal for the contractor to target state s. Furthermore, the thresholds are decreasing in time, i.e. $i^*_1 \geq i^*_2 \geq ... \geq i^*_{n-1}$.*

Theorem 1 states that in every period the contractor either stays at the initial state or performs maintenance to bring the performance to the best target state possible. Furthermore, if the incentive scheme satisfies certain conditions, then the contractor's optimal policy depend on a series of threshold states that are decreasing in time. That is, the contractor is more likely to achieve higher performing pavement states in later periods, given the same initial state.

There are a set of parallel conditions, labeled I5–I8 below, for which the contractor is more likely to target state s in earlier time periods and stay at the initial state later in the life of the structure. We present these conditions below and the result in Theorem 2.

I5: -$\Delta c_t(i,s) \leq \alpha \sum_{0\leq k\leq i} a_{i-k}\{h_n(k+1) - h_n(k)\} \quad t = 1, ..., n-1$

I6: $\Delta^2 h_t(i) - \Delta^2 c_t(i,s) \geq 0 \quad t = 1, ..., n-1$

I7: $\Delta h_{t+1}(i) - \Delta h_t(i) \leq \min\{0, \Delta c_{t+1}(i,s) - \Delta c_t(i,s)\} \quad t = 1, ..., n-2$

I8: $\Delta h_n(i) - \Delta h_{n-1}(i) \leq -\Delta c_{n-1}(i,s)$, $(1-\alpha)\Delta h_n(i) \leq \Delta h_{n-1}(i)$

THEOREM 2 *If the cost of targeting state to s from each fixed initial state is increasing in time and the incentive scheme satisfies inequalities I5–I8, then there exist thresholds i^*_t such that if in period t initial state is below the threshold, the contractor targets state s and if the initial state is at or above the threshold, then it is optimal for the contractor to stay at the initial state. Furthermore, the thresholds are decreasing in time, i.e. $i^*_1 \geq i^*_2 \geq ... \geq i^*_{n-1}$.*

Although there are potentially an infinite number of I/D schemes that satisfy the above conditions we present simple examples below involving linear I/D schemes and linear costs. These results are presented as Corollaries to the Theorems presented above.

COROLLARY 1 *If cost of effort and incentives are linear i.e. $c_t(i,j) = c_t \cdot (j-i)$ and $h_t(i) = h_t \cdot (i - \bar{i}_t)$ for all $t = 1, ..., n-1$, $c_1 \geq c_2 ... \geq c_{n-1}$, and if*

- $h_1 \leq h_2 ... \leq h_{n-1}$,
- $h_{t+1} - h_t \geq (c_t - c_{t+1})$ for all $t = 1, ..., n-2$ and $h_n - h_{n-1} \geq c_{n-1}$,
- $(1-\alpha)h_n \geq h_{n-1}$ *and,*
- $c_{n-1} \geq \alpha h_n \sum_{0\leq k\leq s} a_{s-k}$

then the thresholds are non-increasing in t. That is, $i^*_1 \geq i^*_2 ... \geq i^*_t \geq i^*_{t+1} \geq ... \geq i^*_{n-1}$.

Corollary 1 states that if the cost for unit improvement in state is decreasing in time, then by choosing the I/D scheme such that the incentive for unit improvement increases with time and the rate of increase is at least as large as the rate of decrease of costs, the agency can ensure that the optimal thresholds decrease in time. Because the initial thresholds are higher, the contractor is not likely to exert effort early in the life of the structure. In later periods, the thresholds are lower and so the contractor is more likely to improve pavement state.

COROLLARY 2 *If cost of effort and incentives are linear i.e.* $c_t(i,j) = c_t \cdot (j-i)$ *and* $h_t(i) = h_t \cdot (i - \bar{i_t})$ *for all* $t = 1, ..., n-1$, $c_1 \leq c_2 ... \leq c_{n-1}$, *and if*

- $h_1 \geq h_2 ... \geq h_{n-1}$,
- $h_{t+1} - h_t \leq (c_t - c_{t+1})$ *for all* $t = 1, ..., n-2$ *and* $h_n - h_{n-1} \leq c_{n-1}$,
- $(1-\alpha)h_n \leq h_{n-1}$ *and,*
- $c_{n-1} \leq \alpha h_n a_0$,

then the thresholds are non-increasing in t*. That is,* $i_1^* \geq i_2^* ... \geq i_t^* \geq i_{t+1}^* \geq ... \geq i_{n-1}^*$.

Corollary 2 states that if the cost for unit improvement in state is increasing in time, then by choosing the I/D parameters to be decreasing in time, it is possible to get a series of thresholds that decrease with time. This means that the contractor is more likely to perform maintenance activities early in the life of the structure, and not exert effort in later periods.

For more general cost structures, it is difficult to obtain meaningful expressions for the values of thresholds. However, the formulation introduced above can be used to numerically find precise thresholds for every period for each known effort cost and reward function. Upon knowing $\Pi_1(i)$ for all states i, and the state realized at the end of construction in period 0 (denoted by i_0), the contractor is able to compute its net profit for any realized state i_0, which we denote by $\Pi_0(i_0) = h_0(i_0) + \sum_{k \leq i_0} P_{i_0,k}\Pi_1(k)$. It is also possible to find the optimal target state in period 0 (i.e. the quality of the build) by choosing the profit maximizing state.

5.4 Bidding Strategy

Suppose that before bidding, each contractor receives a signal that allows it to calculate its construction and maintenance effort costs, which are assumed to be private and independent, for each realization of the pavement state. Then, although uncertainty in pavement state is not fully resolved at the time of bidding, each contractor can ex ante calculate its expected profit. Let x denote an arbitrary contractor's minimum bid, i.e. a bid amount that makes its expected profit equal to zero. Clearly, x is the contractor's minimum expected cost and hereafter, we assume that each contractor can calculate x. At the time of bidding, there is uncertainty in cost, but that uncertainty is due to the process of deterioration, which is assumed to be exogenous and common knowledge. Each contractor has the same information about the uncertainty in deterioration. Therefore, following standard arguments developed for first-price sealed bid auctions, it can be established that each contractor will bid an amount equal to the expected value of the second lowest bid, conditioned upon its own bid being the smallest among all bids. That is, each contractor will bid the conditional expected value of the second lowest order statistics given that its own observed minimum bid is the smallest among all bids (details can be found in [23]).

Specifically, consider an arbitrary contractor indexed 1 and let there be ν contractors. Then, Contractor 1 wins the auction only if all other contractors have higher cost. Denote by $Y_1^{(\nu-1)}(x_1)$ the second lowest cost among the contractors, conditioned upon Contractor 1 having the lowest cost x_1. That is $x_1 \leq x_m$ for each $m \in \{1, \cdots, \nu\}$. Then, Contractor 1 bids $E[Y_1^{(\nu-1)}(x_1)]$. The extra payment made by the agency to the contractors is factored in its bid because of competitive pressures, as a contractor expecting to receive an incentive will include that amount into its calculation for minimum expected cost x. The expected cost of the agency is the overall expected cost of the second highest minimum cost among all contractors. That is, the agency does not pay more to contractors over and above the cost of achieving the desired pavement state. The extra amount earned by the lowest cost bidder is determined by the number of bidders and competitive advantage of some bidders over others.

Our analysis so far assumed that the agency knows the effort costs functions faced by contractors. In reality, agency's estimates are uncertain. In that case, Theorem 1 and Corollary 1 (resp. Theorem 2 and Corollary 2) will hold if sets I1–I4 (resp. I5–I8) are satisfied for all contractors. We realize that this may not be possible and when that happens it is possible for some contractors to earn windfall profits on account of informational asymmetry. This happens because the I/D scheme must be specified before a request for bids is issued and the same schedule must be applied to all contractors, regardless of who wins the bid.

5.5 Implementation Issues

In practice, STAs must consider additional issues that arise during implementation that are not captured in the basic model presented above.

The measurement of performance and payment of incentive by the STA, as well as both the decision and the implementation of the maintenance procedure by the contractor all occur simultaneously in the model. Contractors have an incentive to do maintenance immediately before performance measurement, which may conflict with the STA's goal of high performance at all times. The timing of inspections varies from STA to STA, and MnDOT specifies inspection methods in [22]. An approach to minimize the gaming of the system by the timing of measurements is to randomize (within a range) the inspection time.

Since the act of maintenance itself is disruptive to road users, a method to provide a general disincentive to disruptive maintenance is necessary to properly align the incentives of the STA and the contractor. Lane rental provisions, where the contractor must pay a penalty for preventing road usage, satisfies this goal without changing the overall results, as the additional penalties are reflected in the bids the same way that incentives are.

Contracts must be responsive to effects that are outside of the contractor's control. For instance, if traffic loading changes significantly, or temperature fluctuates beyond expectations at the time the contract was executed, adjustments to performance specifications should be allowed. The thresholds at which the deviations change the specifications should be determined in the initial contract, as should the method by which the specifications or incentives will change.

The STA must also be prepared for the possibility that the contractor's financial stability cannot be guaranteed for the duration of the contract. Since payments between the two parties are ongoing throughout the contract period, if the contractor is unable to either provide the necessary maintenance or to pay the assessed penalties, the STA must have contingencies. A high-risk contractor may be incentivized to bid lower than a more stable contractor with an identical cost, as its losses are essentially bounded by bankruptcy. One approach, likely controversial, is to assess a premium (through smaller incentives) to contractors unable to demonstrate their financial stability. Another is to require all contractors to share the cost of potential failed contractors through some form of risk-sharing mechanism (e.g., all contractors must buy some form of insurance). However, the details and feasibility of such a system are beyond the scope of this report.

6 Conclusions

We discussed various contracts that have been used in industry and argued that PBCs can achieve better coordination than existing contracting methods. PBCs more accurately reward/penalize the contractor because they are based on in-use performance, not predicted performance, as in the case of PBS. PBCs encourage contractors to innovate and perform timely preventive maintenance activity, improving the long-term performance of the structure. However, implementation of PBCs is challenging because of more contract parameters and more supervisory/measurement effort.

We noted that distress prediction can be done quickly using CFS. CFS for rutting and fatigue cracking are expressed as a function of effective dynamic modulus (E^*) that incorporates mix, traffic and climate properties. From the CFS, we inferred that predicted distress follows the law of diminishing returns. In addition to the dynamic modulus, the contractor can also control the timing and extent of preventive maintenance activities to affect pavement performance. Reducing variability in construction practices also goes a long way toward increasing quality and pavement life. Knowing that the contractor can exert effort to significantly improve pavement performance, PBCs can be used to provide the right incentives to contractors. We listed an inventory of performance metrics that STAs can use to evaluate contractor performance.

We then considered how knowledge of distress prediction and contractor maintenance options could be used to design effective PBCs. We showed that under reasonable assumptions about cost functions, it would be optimum for the contractor to either stay at the initial state or exert effort to bring the pavement to the best possible state. The contractor's optimal action in each period depended on where the initial state lay relative to a threshold for that period and whether the costs were increasing or decreasing over time. Our analysis confirmed that it would be possible for the agency to choose incentives to elicit a desired response from the contractor.

We demonstrated this approach when the cost of effort and incentives are linear. In that case, it is possible to derive an ordering of the thresholds. If the cost for unit improvement in state is decreasing in time, then by choosing the I/D scheme such that the incentive for unit improvement increases with time and the rate of increase is at least as large as the rate of decrease of costs, the optimal thresholds of the contractor decrease in time. Under this scheme, the initial thresholds are higher and the contractor is more likely to not exert effort early in the life of the structure. In later years, the thresholds are lower and so the contractor is more likely to perform maintenance activity to maintain the best state. Similarly if the costs for unit improvement are increasing over time, providing decreasing incentives over time leads to decreasing thresholds that cause the contractor to perform more maintenance activities early in the life of the structure. That is, the choice of incentive scheme can be matched with the known structure of maintenance costs to realize the desired contractor effort on maintaining pavement state.

We used standard arguments from procurement auctions to argue that in expectation, when using PBCs as an alternative to other methods, agencies pay only for improved quality and that contractors do not earn unreasonable excess profits. We then discussed some implementation issues that any efforts to implement PBCs must consider.

References

[1] G. P. Cachon, "Supply Chain Coordination with Contracts," in *Handbooks in Operations Research and Management Science: Supply Chain Management*, edited by G. Steve and T. De Kok, North-Holland, 2003.

[2] D. Murty and I. Djamaludin, "New product warranty: A literature review," *International Journal of Production Economics*, **79**, 3, pp. 231-236, 2002.

[3] R. J. Gilbertson, "Risk reduction through application of essential performance requirements warranties," in *Proceedings of the IEEE Aerospace and Electronics Conference*, Dayton, OH, pp. 2010-2014, 1989.

[4] I. K. Battaglia and S. W. Krebs, "Twelve-Year Evaluation of the Wisconsin Pavement Warranty Program," in *Proceedings of the Transportation Research Board 89th Annual Meeting*, Washington, D.C., 2010.

[5] C. J. Wienrank, *Demonstrating the use of Performance-Based Warranties on Highway Construction Projects*, Illinois Department of Transportation, Springfield, IL, 2004.

[6] M. Lu and C. Donaldson, "Performance-based contracts and provider efficiency: The state of the art," *Disease Management and Health Outcomes*, **7**, 3, pp. 127-137, 2000.

[7] T. Stenbeck, "Quantifying effects of incentives in a rail maintenance performance-based contract," *Journal of Construction Engineering and Management*, **134**, 4, pp. 265-272, 2008.

[8] Pollution Engineering, "Ahead of Schedule, Under Budget." URL: http://www.pollutionengineering.com/Archives?issue=132661, 2006. Downloaded: 27 Mar 2010.

[9] R. Cerar, "US Army Installation Restoration Program Performance-Based Contracts," *Engineer Professional Bulletin*, **6**, 4, pp. 38-39. URL: http://www.wood.army.mil/engrmag/PDFs%20for%20Oct-Dec%2004/Cerar.pdf, 2004. Downloaded: 20 Mar 2010.

[10] B. L. Bateman and J. C. Wetherbe, "Cost analysis of computer maintenance contracts," *MIS Quarterly*, **2**, 4, pp. 15-22, 1978.

[11] W. A. Hyman, *NHCRP Synthesis 389: Performance-based contracting for maintenance*, National Cooperative Highway Research Program, 2009.

[12] Partnerships British Columbia, *Achieving Value for Money: Sea-to-Sky Highway Improvement Project*, British Columbia Ministry of Transportation. URL: http://www.partnershipsbc.ca/pdf/SeatoSkyFinal.pdf, 2006. Downloaded: 8 May 2010.

[13] Federal Highway Administration, *Michigan Demonstration Project: Performance Contracting for Construction on M-115 in Clare County, MI*. URL: http://www.fhwa.dot.gov/hfl/summary/mi/, 2009. Downloaded: 21 Apr 2011.

[14] M. El-Basyouny and M. G. Jeong, "Development of Database Solution for Prediction of the MEPDG Permanent Deformation," in *Proceedings of the Transportation Research Board 89th Annual Meeting*, Washington, D.C., 2010.

[15] M. El-Basyouny and M. G. Jeong, "Development of Probabilistic Performance Related Specifications Methodology Based upon AASHTO MEPDG," in *Proceedings of the Transportation Research Board 89th Annual Meeting*, Washington, D.C., 2010.

[16] M. G. Jeong, J. Uzan, and M. W. Witczak, "Development of the Stochastic Performance Related Specification Framework for Thermal Cracking Distress of Asphalt Pavement Using the AASHTO MEPDG Prediction Models," in *Proceedings of the Transportation Research Board 89th Annual Meeting*, Washington, D.C., 2010.

[17] S. M. El-Badawy, M. G. Jeong, and M. El-Basyouny, "Methodology to predict alligator fatigue cracking distress based on asphalt concrete dynamic modulus," *Transportation Research Record*, 2095, pp. 115-126, 2010.

[18] K. Hoegh, L. Khazanovich, and M. Jensen, "Local calibration of MEPDG rutting model for MnRoad test sections," in *Proceedings of the Transportation Research Board 89th Annual Meeting*, Washington, D.C., 2010.

[19] M. Marasteanu, A. Zofka, M. Turos, X. Li, R. Velasquez, X. Li, W. Buttlar, G. Paulino, A. Braham, E. Dave, J. Ojo, H. Bahia, C. Williams, J. Bausano, A. Gallistel, and J. McGraw, *Investigation of low temperature cracking in asphalt pavements, National Pooled Fund Study 776*, Technical Report MN/RC 2007-43, MnDOT. URL: http://www.cts.umn.edu/Publications/ResearchReports/reportdetail.html?id=1522, 2007. Downloaded: 21 Apr 2011.

[20] G. P. Ong, T. E. Nantung, and K. C. Sinha, "An Assessment of Procedures to Determine Intervention Levels for Pavement Preservation," in *Proceedings of the Transportation Research Board 89th Annual Meeting*, Washington, D.C., 2010.

[21] W. S. Guthrie, T. Rogers, and P. M. Ellis, "Variability in Construction of Cement-Treated Base Layers: Probabilistic Analysis of Pavement Life Using Mechanistic-Empirical Approach," in *Proceedings of the Transportation Research Board 89th Annual Meeting*, Washington, D.C., 2010.

[22] MnDOT, "An Overview of Mn/DOT's Pavement Condition Rating Procedures and Indices." URL: http://www.dot.state.mn.us/materials/Rating_Overview_State.pdf, 2006. Downloaded: 27 Mar 2010.

[23] V. Krishna, *Auction Theory*, Academic Press, Elsevier, San Diego, CA, 2002.

Appendix A

Technical Proofs

We prove Theorem 1 below and omit the proof of Theorem 2, which is similar to the proof of Theorem 1. The latter is presented in two parts. In Part 1, we show that optimal contractor action in every period is either to stay at the initial state i or to target state s and that there exists a threshold state in every period such that if the initial state is below the threshold, the contractor stays at the initial state and if the initial state is at or above the threshold, then the contractor achieves state s. In Part 2, we show that the thresholds are decreasing in time.

Proof of Theorem 1:

Part 1

Recall that $\Delta f(i) = f(i+1) - f(i)$ and $\Delta^2 f(i) = \Delta f(i+1) - \Delta f(i)$, where f is an arbitrary function defined over discrete i. If f has two arguments i and j, then we define $\Delta f(i,j) = f(i,j+1) - f(i,j)$. We first have,

$$\begin{aligned}
\Delta g_t(i,j+1) &= -\Delta c_t(i,j+1) + \alpha \sum_{0\le k\le j+2} P_{j+2,k}\Pi_{t+1}(k) - \alpha \sum_{0\le k\le j+1} P_{j+1,k}\Pi_{t+1}(k) \\
&= -\Delta c_t(i,j+1) + \alpha \sum_{0\le k\le j+1} a_{j+1-k}[\Pi_{t+1}(k+1) - \Pi_{t+1}(k)]
\end{aligned}$$

where

$$P_{j,k} = \begin{cases} P(A = j-k) = a_{j-k} & \text{if } k = j, j-1, ..., 1, \\ P(A \ge j) = \sum_{r\ge j} a_r & \text{if } k = 0, \end{cases}$$

and A $\in \mathcal{I}^+$. Notation $\mathcal{I}^+$ represents the set of nonnegative integers.

Using similar arguments we can show that,

$$\begin{aligned}
\Delta g_t(i,j) &= -\Delta c_t(i,j) + \alpha \sum_{0\le k\le j+1} P_{j+1,k}\Pi_{t+1}(k) - \alpha \sum_{0\le k\le j} P_{j,k}\Pi_{t+1}(k) \\
&= -\Delta c_t(i,j) + \alpha \sum_{0\le k\le j} a_{j-k}[\Pi_{t+1}(k+1) - \Pi_{t+1}(k)]
\end{aligned}$$

From above we can simplify $\Delta g_t(i,j+1) - \Delta g_t(i,j)$ as follows. Note this is defined only for $j \ge i$.

$$\Delta g_t(i,j+1) - \Delta g_t(i,j) = \begin{array}{l} -[\Delta c_t(i,j+1) - \Delta c_t(i,j)] + \alpha a_{j+1}[\Pi_{t+1}(1) - \Pi_{t+1}(0)] \\ +\alpha \sum_{k=0}^{j} \Big\{[\Pi_{t+1}(k+2) - \Pi_{t+1}(k+1)] - [\Pi_{t+1}(k+1) - \Pi_{t+1}(k)]\Big\} \end{array}$$

Because we have assumed that rate of increase of cost is decreasing, we have that $\Delta c_t(i,j+1) - \Delta c_t(i,j)] \le 0$ for all $j \ge i$. Suppose, by induction hypothesis, $\Pi_{t+1}(.)$ is increasing convex in its argument and that the rate of increase of the profit functions is increasing in time for decision epochs at or after $t+1$. That is, we assume that

- $[\Pi_{t+1}(1) - \Pi_{t+1}(0)] \ge 0$,
- $\{[\Pi_{t+1}(k+2) - \Pi_{t+1}(k+1)] - [\Pi_{t+1}(k+1) - \Pi_{t+1}(k)]\} \ge 0$, and

- $[\Pi_l(i+1) - \Pi_l(i)]$ is increasing in l for each i and $l \geq t+1$.

Then by the first and second induction assumptions we have $\Delta g_t(i, j+1) - \Delta g_t(i, j) \geq 0$. That is, the rate of increase of $g_t(i, j)$ is increasing in j for each fixed i. Therefore, the contractor's optimal action is period t is to either stay at the initial state or target state s.

Next, we argue that the contractor's optimal action in period t is monotone in the initial state. To see this we find conditions such that $[g_t(i+1, s) - g_t(i+1, i+1)] - [g_t(i, s) - g_t(i, i)] \geq 0$ for all i. Note that,

$$\begin{aligned}\Delta[g_t(i,s) - g_t(i,i)] &= \Big(c_t(i,s) + \alpha \sum_{0 \leq k \leq i} P_{i,k}\Pi_{t+1}(k)\Big) - \Big(c_t(i+1,s) + \alpha \sum_{0 \leq k \leq i+1} P_{i+1,k}\Pi_{t+1}(k)\Big) \\ &= c_t(i,s) - c_t(i+1,s) - \alpha \sum_{0 \leq k \leq i} a_{i-k}[\Pi_{t+1}(k+1) - \Pi_{t+1}(k)].\end{aligned}$$

By induction hypothesis we have that $\Pi_{t+1}(k+1) - \Pi_{t+1}(k) \leq \Pi_n(k+1) - \Pi_n(k) = h_n(k+1) - h_n(k)$.

Thus, a sufficient condition for $[g_t(i+1, s) - g_t(i+1, i+1)] - [g_t(i, s) - g_t(i, i)] \geq 0$ is that $-c_t(i+1, s) + c_t(i, s) \geq \alpha \sum_{0 \leq k \leq i} a_{i-k}[h_n(k+1) - h_n(k)]$ for all $t = 1, ..., n-1$. This means that when the decrease in cost on account of targeting state s from initial state $(i+1)$ versus initial state i is greater than or equal to the expected increase in incentive payment from improving state by 1 in period n, then the optimal actions are increasing in the initial state.

The above arguments imply that there must exist a threshold i_t^* such that if the initial state is less than the threshold, the contractor's optimal action is to do nothing and if the initial state is greater than or equal to the threshold, the contractor brings the final state to s. The formal definition of i_t^* is as follows:

$$i_t^* = \{i \in 0, 1, ..., s | g_t(i-1, s) - g_t(i-1, i-1) < 0, g_t(i, s) - g_t(i, i) \geq 0\}.$$

To complete the induction we need to show that if the induction hypothesis for period $t+1$ were true, then it would true in period t as well. That is we need to show that,

- $[\Pi_t(1) - \Pi_t(0)] \geq 0$
- $\{[\Pi_t(k+2) - \Pi_t(k+1)] - [\Pi_t(k+1) - \Pi_t(k)]\} \geq 0$
- $[\Pi_l(i+1) - \Pi_l(i)]$ is increasing in l for each i and $l \geq t$

We first show that $\Pi_t(1) - \Pi_t(0) \geq 0$. The optimal action when the initial state is i can be either to stay at the initial state i or target state s. Suppose the initial state is 0 and the contractor's optimal action is target s in period t. Then, it is also optimal for the contractor to target state s when the initial state is 1. If however the contractor chooses to stay at the initial state 0, there can be two possibilities in regards to its optimal action if the initial state were 1 — either stay at state 1 or target s. These arguments result in three cases that are analyzed below.

Case 1: $j_t^*(0) = s$, $j_t^*(1) = s$

$\Pi_t(1) - \Pi_t(0) = h_t(1) - h_t(0) + g_t(1, s) - g_t(0, s)$

$= h_t(1) - h_t(0) - c_t(1, s) + c_t(0, s) \geq 0$

Case 2: $j_t^*(0) = 0$, $j_t^*(1) = 1$
$\Pi_t(1) - \Pi_t(0) = h_t(1) - h_t(0) + g_t(1,1) - g_t(0,0)$
$= h_t(1) - h_t(0) + \alpha \sum_{0 \le k \le 1} P_{1,k}\Pi_{t+1}(k) - \alpha \sum_{0 \le k \le 0} P_{0,k}\Pi_{t+1}(k)$
$= h_t(1) - h_t(0) + \alpha a_0[\Pi_{t+1}(1) - \Pi_{t+1}(0)] \ge 0$

Case 3: $j_t^*(0) = 0$, $j_t^*(1) = s$
$\Pi_t(1) - \Pi_t(0) = h_t(1) - h_t(0) + g_t(1,s) - g_t(0,0)$
$\ge h_t(1) - h_t(0) + g_t(1,1) - g_t(0,0) \ge 0$

Thus, $\Pi_t(1) - \Pi_t(0) \ge 0$.

Next we show that $\{[\Pi_t(i+2) - \Pi_t(i+1)] - [\Pi_t(i+1) - \Pi_t(i)]\} \ge 0$. If when the initial state is i the optimal action is to target s, then that must be optimal action when the initial state is $i+1$ and $i+2$ as well. If, instead, the optimal action is to do nothing, then we consider the two choices that the contractor would have if it were in state $i+1$. If the contractor targets s from the initial state $i+1$, then it will do so for an initial state of $i+2$ as well. If however, the contractor's optimal action is to stay at the initial state $i+1$, then again there are two choices in period $i+2$. That is, we can focus on the following four cases:

1. $j_t^*(i) = s$, $j_t^*(i+1) = s$, $j_t^*(i+2) = s$
2. $j_t^*(i) = i$, $j_t^*(i+1) = s$, $j_t^*(i+2) = s$
3. $j_t^*(i) = i$, $j_t^*(i+1) = i+1$, $j_t^*(i+2) = s$
4. $j_t^*(i) = i$, $j_t^*(i+1) = i+1$, $j_t^*(i+2) = i+2$

We analyze each case separately.

Case 1: $j_t^*(i) = s$, $j_t^*(i+1) = s$, $j_t^*(i+2) = s$
$[\Pi_t(i+2) - \Pi_t(i+1)] - [\Pi_t(i+1) - \Pi_t(i)] = \Delta^2\Pi_t(i) = \Delta^2 h_t(i) - \Delta^2 c_t(i,s)$

Thus, the above expression is non negative iff $\Delta^2 h_t(i) - \Delta^2 c_t(i,s) \ge 0$. This means that the rate of increase of the incremental differences in the reward must be greater than or equal to the rate of decrease of incremental differences in the cost when the target state is s.

Case 2: $j_t^*(i) = i$, $j_t^*(i+1) = s$, $j_t^*(i+2) = s$
$\Delta^2\Pi_t(i) = \Delta^2 h_t(i) + [g_t(i+2,s) - g_t(i+1,s)] - [g_t(i+1,s) - g_t(i,i)]$
$\ge \Delta^2 h_t(i) + [g_t(i+2,s) - g_t(i+1,s)] - [g_t(i+1,s) - g_t(i,s)] \ge 0$

Case 3: $j_t^*(i) = i$, $j_t^*(i+1) = i+1$, $j_t^*(i+2) = i+2$
$\Delta^2\Pi_t(i) = \Delta^2 h_t(i) + [g_t(i+2,i+2) - g_t(i+1,i+1)] - [g_t(i+1,i+1) - g_t(i,i)]$
$= \Delta^2 h_t(i) + \alpha \sum_{1 \le k \le i} a_{i-k} \cdot \{\Delta^2\Pi_{t+1}(k)\} + \alpha a_i\{\Pi_{t+1}(1) - \Pi_{t+1}(0)\} \ge 0$

Case 4: $j_t^*(i) = i$, $j_t^*(i+1) = i+1$, $j_t^*(i+2) = s$
$\Delta^2\Pi_t(i) = \Delta^2 h_t(i) + [g_t(i+2,s) - g_t(i+1,i+1)] - [g_t(i+1,i+1) - g_t(i,i)]$
$\ge \Delta^2 h_t(i) + [g_t(i+2,i+2) - g_t(i+1,i+1)] - [g_t(i+1,i+1) - g_t(i,i)] \ge 0$

This completes the proof that $\{[\Pi_t(i+2)-\Pi_t(i+1)]-[\Pi_t(i+1)-\Pi_t(i)]\} \geq 0$.

Next, we show that $\Delta\Pi_t(i) \leq \Delta\Pi_{t+1}(i)$. Combined with the induction assumption that $\Pi_l(i+1)-\Pi_l(i)$ is non-decreasing in l for all $l \geq t+1$, we would have $\Pi_l(i+1)-\Pi_l(i)$ is non-increasing in i for all $l \geq t$, completing the induction. Note that,

$$\begin{aligned}
&\Pi_t(i+1)-\Pi_t(i) \leq \Pi_{t+1}(i+1)-\Pi_{t+1}(i) \\
\Longleftrightarrow\ &\Pi_{t+1}(i)-\Pi_t(i) \leq \Pi_{t+1}(i+1)-\Pi_t(i+1)
\end{aligned}$$

Thus it suffices to show that $\Pi_{t+1}(i)-\Pi_t(i)$ is increasing in i. Here we have two possibilities for the contractor's optimal action in period t for an initial state i — either stay at i or target s. Similarly the contractor has the same two choices in period $t+1$, resulting in the following four cases:

Case 1: $j_t^*(i) = i$, $j_{t+1}^*(i) = i$
$\Pi_{t+1}(i)-\Pi_t(i) = h_{t+1}(i)$ - $h_t(i)$ + $g_{t+1}(i,i)$ - $g_t(i,i)$
= $h_{t+1}(i)$ - $h_t(i)$ + $\alpha\sum_{0\leq k\leq i} P_{i,k}\Big(\Pi_{t+2}(k)-\Pi_{t+1}(k)\Big)$

We assume that $h_{t+1}(i)$ - $h_t(i)$ is increasing in i and claim that $\alpha\sum_{0\leq k\leq i} P_{i,k}\Big(\Pi_{t+1}(k)-\Pi_t(k)\Big)$ in increasing in i since $\Big(\Pi_{t+2}(k)-\Pi_{t+1}(k)\Big)$ is increasing in k for all $0 \leq k \leq i$. [Note: $\Pi_{t+2}(k)-\Pi_{t+1}(k)$ is increasing in k because $\Delta\Pi_{t+2}(k) \geq \Delta\Pi_{t+1}(k)$ by induction assumption]. We prove this in a lemma below.

Lemma 1:
$h(i) = \sum_{0\leq k\leq i} P_{i,k} f(k)$ is increasing(decreasing) in i, if $f(k)$ is increasing(decreasing) in k for all $0 \leq k \leq i$

Proof:
Note that,
$h(i+1)-h(i) = \alpha\sum_{0\leq k\leq i+1} P_{i+1,k} f(k) - \alpha\sum_{0\leq k\leq i} P_{i,k} f(k) = \alpha\sum_{0\leq k\leq i} a_{i-k}\Big(f(k+1)-f(k)\Big)$
from which the above result follows. #

Case 2: $j_t^*(i) = s$, $j_{t+1}^*(i) = s$
$\Pi_{t+1}(i)-\Pi_t(i) = h_{t+1}(i)$ - $h_t(i)$ + $g_{t+1}(i,s)$ - $g_t(i,s)$
= $h_{t+1}(i)$ - $h_t(i)$ + $c_t(i,s)-c_{t+1}(i,s)$ +$\alpha\sum_{0\leq k\leq s} P_{s,k}\Big(\Pi_{t+2}(k)-\Pi_{t+1}(k)\Big)$

If $h_{t+1}(i)$ - $h_t(i)$ + $c_t(i,s)-c_{t+1}(i,s)$ were increasing in i, then using Lemma 1, $\Pi_{t+1}(i)-\Pi_t(i)$ is increasing in i. Thus a sufficient condition is that $\Delta h_{t+1}(i)-\Delta h_t(i) \geq \Delta c_{t+1}(i,s)-\Delta c_t(i,s)$.

Case 3: $j_t^*(i) = i$, $j_{t+1}^*(i) = s$
$\Pi_{t+1}(i)-\Pi_t(i) = h_{t+1}(i)$ - $h_t(i)$ + $g_{t+1}(i,s)$ - $g_t(i,i)$
$\geq h_{t+1}(i)$ - $h_t(i)$ + $g_{t+1}(i,i)$ - $g_t(i,i)$

Since the right hand side (RHS) of the above inequality in increasing in i the left hand side (LHS) must also be increasing in i.

Case 4: $j_t^*(i) = s$, $j_{t+1}^*(i) = i$
$\Pi_{t+1}(i) - \Pi_t(i) = h_{t+1}(i) - h_t(i) + g_{t+1}(i,i) - g_t(i,s)$
$\geq h_{t+1}(i) - h_t(i) + g_{t+1}(i,s) - g_t(i,s)$

Since the RHS of the above inequality in increasing in i the LHS must also be increasing in i.

Thus, we see that $\Pi_{t+1}(i) - \Pi_t(i)$ is non-decreasing in i and so $\Pi_l(i+1) - \Pi_l(i)$ is increasing in l for all $l \geq t$. To complete the induction we need to ensure that the claims hold true in period $n-1$ as well. That is, we want,

- $\Pi_n(1) - \Pi_n(0) = h_n(1) - h_n(0) \geq 0$
- $\{[h_n(k+2) - h_n(k+1)] - [h_n(k+1) - h_n(k)]\} \geq 0$
- $h_n(i) - \Pi_{n-1}(i)$ is increasing in i

The first two are true by assumption. It can be shown, similar to the earlier arguments, that the conditions sufficient to ensure that $h_n(i) - \Pi_{n-1}(i)$ is increasing in i are that, (a) $\Delta h_n(i) - \Delta h_{n-1}(i) \geq -\Delta c_{n-1}(i,s)$, and (b) $(1-\alpha)\Delta h_n(i) \geq \Delta h_{n-1}(i)$ for all i.

In summary, by using induction for all periods we have that the optimal contractor action in every period is either to stay at the initial state i or bring the final state to s and that there exists a threshold state in every period such that if initial state is below the threshold, then the contractor stays at the initial state and if the initial state is at or above the threshold, then it is optimal for the contractor to target s.

Part 2
In Part 2, we prove that the thresholds are indeed decreasing with time. To do this we assume that,

- $i_{t+1}^* \geq i_{t+2}^* \geq ... \geq i_{n-1}^*$ where,
  $i_t^* = \{i \in 0, 1, ..., s | g_t(i-1,s) - g_t(i-1,i-1) < 0, g_t(i,s) - g_t(i,i) \geq 0\}$

and show that $i_t^* \geq i_{t+1}^*$ which will complete the induction. We have,

$\Big(g_{t+1}(i,s) - g_{t+1}(i,i)\Big) - \Big(g_t(i,s) - g_t(i,i)\Big) =$
$c_t(i,s) - c_{t+1}(i,s) + \alpha \sum_{0 \leq k \leq s} P_{s,k}\Big(\Pi_{t+2}(k) - \Pi_{t+1}(k)\Big) - \alpha \sum_{0 \leq k \leq i} P_{i,k}\Big(\Pi_{t+2}(k) - \Pi_{t+1}(k)\Big)$

By assumption $c_t(i,s) - c_{t+1}(i,s) \geq 0$.

Also $\alpha \sum_{0 \leq k \leq s} P_{s,k}\Big(\Pi_{t+2}(k) - \Pi_{t+1}(k)\Big) \geq \alpha \sum_{0 \leq k \leq i} P_{i,k}\Big(\Pi_{t+2}(k) - \Pi_{t+1}(k)\Big)$
since $\Pi_{t+2}(k) - \Pi_{t+1}(k)$ is increasing in k for all k.

Thus, $\Big(g_{t+1}(i,s) - g_{t+1}(i,i)\Big) - \Big(g_t(i,s) - g_t(i,i)\Big) \geq 0$ and so $i_t^* \geq i_{t+1}^*$
(by definition of i_t^*)

By induction $i_t^* \geq i_{t+1}^* \geq ... \geq i_{n-1}^*$. Thus we have shown that if the thresholds are non-increasing period $t+1$ onwards then they are non-increasing from period t to $t+1$ as well. By induction the thresholds are non-increasing for all t.

Proofs of Corollaries

In the corollaries, all expressions are obtained by substituting the linear expressions for the incentives and costs in the corresponding expressions in the Theorems. One expression of note in Corollary 1 is $c_t \geq \alpha h_n \sum_{0 \leq k \leq i} a_{i-k}$ for all i and t. Since the costs are decreasing in time a sufficient condition to ensure the above for all i and t is to let, $c_{n-1} \geq \alpha h_n \sum_{0 \leq k \leq s} a_{s-k}$.
Similarly in Corollary 2, having $c_{n-1} \leq \alpha h_n a_0$ is sufficient to ensure that $c_t \leq \alpha h_n \sum_{0 \leq k \leq i} a_{i-k}$ for all i and t.