ESTIMATION OF LATENT PAVEMENT PERFORMANCE FROM DAMAGE MEASUREMENTS

by

ROHIT RAMASWAMY

Submitted to the Department of Civil Engineering on June 29, 1989 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Civil Engineering

ABSTRACT

A pavement deterioration model predicts the performance of a pavement over time as a function of traffic, pavement characteristics and environmental factors. The most important performance characteristics of a pavement are its ability to bear traffic loads and its ability to provide a smooth ride. However, there is no unambiguous approach to directly measure these performance characteristics. Therefore, we consider pavement performance to be unobservable. The problem of designing pavement deterioration models is the problem of defining the above unobservable characteristics in terms of what is observed, i.e., in terms of measured extents and severities of different damage components. One approach that has been applied is to regress subjective judgements of inspectors evaluating the pavement on a prespecified set of damage measurements. The two most widely used indices of pavement performance—the PCI and the PSI have been derived in this fashion on the basis of engineering judgement. A method of getting around the reliance on a subjective condition index has been to model the deterioration over time of a small set of observed damage components.

The methodology presented in this thesis is a statistically rigorous technique that both avoids judgemental measures and provides a basis for a choice between different damage measurements and alternative damage models. Pavement performance measures are statistically estimated from observed pavement damage. No constraints are placed on the number or type of measurements required, so the methodology is flexible enough to include different measurement techniques and data collection strategies. Choice between different measurements is made on the basis of the contribution of a particular measurement to the estimation of latent pavement performance. The estimation procedure simultaneously fits a deterioration model and a performance index calibration model to data, thereby producing much better fits to data than traditional deterioration models.

Another major problem in estimating deterioration models from data collected from in-service pavements is that they do not take into account the fact that observed damage is the result of two simultaneously occurring processes: a process of deterioration and the maintenance performed by the highway agency on the pavement. Even if a model using the PSI as the measure of pavement performance were developed, a proper specification should include these simultaneous effects. In this thesis, a deterioration model is first estimated using the PSI as the performance index and a simultaneous equation specification. This model is then compared against a model using a latent performance variable, where the latent variable is constructed from the same damage measurements as the PSI. Subsequently, models involving multiple latent variables are developed. The methodology presented in this thesis will be useful for
deriving more realistic predictive models of pavement deterioration and for defining better data collection strategies. This methodology is also applicable to other facilities that are subject to deterioration or whose deterioration cannot be easily observed. Examples include bridges, rails and buildings.

Thesis Supervisor: Dr. Moshe Ben-Akiva

Title: Professor of Civil Engineering
ACKNOWLEDGEMENTS

The people whom I would like to thank in this section fall into two categories. The first category consists of those who have contributed to the technical aspects of this dissertation by lending their time and professional expertise. The second category, which is perhaps more important than the first one, consists of those who, through their friendship, have acted as a constant source of emotional sustenance and a life support system over the past five years. The two categories are not rigid; some of those mentioned in one category are deservedly entitled to right-of-place in the other.

In the first category, I would like to thank my thesis supervisor, Prof. Moshe Ben-Akiva for his total accessibility and the significant technical advice and assistance and good ideas that he provided in the course of this dissertation. Discussions with him almost always helped to clarify issues and raise important questions. During the course of this dissertation, I have developed a respect for his conceptual insights and analytical reasoning. Voluminous thanks are also due to Prof. Sue McNeil, who in addition to being a very useful research partner, went over and above the call of duty by extending a friendly hand at critical times. Thanks also to the third member of my committee, Prof. Joe Sussman. Some of the research reported in this dissertation was done under the purview of a project entitled "Theory and Methods for Analysis of Infrastructure Performance and Maintenance Strategies" funded by a grant from the Army Research Office to the Center for Construction Research and Education of the Civil Engineering at MIT.

At the head of the second category, a standing ovation to Patty for being an unfailing source of support even at the most desperate of times. A heartfelt thank you for tolerating, with more patience than was warranted, some of my most unpersonable moments. Thanks to those at MIT who helped to make life easier, especially Bruno, Frannie, Samer, Rabi and others with whom I have come into contact during my stay here. A special thank you to Sue Harmon and Paula for many years of friendship.

Finally, I would like to express sincere appreciation for my family, both here and in India, for keeping open, constantly and without sanction, a refuge to which to return should it ever become necessary.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title Page</td>
<td>1</td>
</tr>
<tr>
<td>Abstract</td>
<td>2</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>4</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>5</td>
</tr>
<tr>
<td>List of Tables</td>
<td>8</td>
</tr>
<tr>
<td>List of Figures</td>
<td>10</td>
</tr>
</tbody>
</table>

**CHAPTER 1. INTRODUCTION**

1.1 A Framework for Infrastructure Performance Analysis  12

1.2 The Latent Variable Approach to Deterioration Modeling  18

1.3 Simultaneity of Deterioration and Maintenance  23

1.4 Contributions of Research  26

**CHAPTER 2. PAVEMENT DETERIORATION MODELS - STATE OF THE ART**

2.1 Performance of a Highway Pavement  29

2.2 Collection of Pavement Damage Data  32

2.3 Modeling Pavement Deterioration from Damage Data  36

2.4 A Review of Pavement Deterioration Models  38

2.5 Limitations of Deterioration Models  52
CHAPTER 3. FORMULATION OF LATENT VARIABLE MODELS 56
3.1 Specification of a General Latent Variable Model 56
3.2 Formulation of the Latent Variable Model for Highway Pavements 59
3.3 Some Special Cases of Linear Latent Variable Models 63
3.4 Effects of Different Error Correlation Patterns 66
3.5 Identification of Latent Variable Models 68

CHAPTER 4. ESTIMATION OF LATENT VARIABLE MODELS 86
4.1 Estimation of a Latent Variable Model 86
4.1.1 Covariance Matrix as a Function of Parameters 86
4.1.2 Estimation of Parameters 89
4.2 Extraction of Latent Variable Scores 94
4.3 Goodness-of-fit Tests 97
4.4 Specification Issues 100

CHAPTER 5. DESCRIPTION OF RESULTS 105
5.1 Description of Data 105
5.2 A Single Equation PSI Model 112
5.3 A Simultaneous Equation PSI Model 116
5.4 A PSI Type Latent Variable Model 120
5.5 A Latent Variable Model with Two Latent Variables 130
5.6 Conclusion 144
CHAPTER 6. APPLICATION OF LATENT VARIABLE DETERIORATION MODELS

6.1 Effect of Additional Measurements
6.2 Prediction of Future Pavement Condition

CHAPTER 7. DETERIORATION MODELS WITH DISCRETE ORDINAL INDICATORS

7.1 State Transitions in a Single Time Period
7.2 Some Models of Facility Deterioration
7.3 Multiple Period Transition Probabilities
7.4 Conclusions

CHAPTER 8. CONCLUSIONS

References
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table No.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2-1</td>
<td>Damage Types for Asphalt Pavements</td>
<td>33</td>
</tr>
<tr>
<td>Table 2-2</td>
<td>Damage Types used for PCI Calculation</td>
<td>39</td>
</tr>
<tr>
<td>Table 5-1</td>
<td>Description of Variables in Nevada Data</td>
<td>108</td>
</tr>
<tr>
<td>Table 5-2(a)</td>
<td>Summary Statistics of Variables</td>
<td>110</td>
</tr>
<tr>
<td>Table 5-2(b)</td>
<td>Kurtosis of Selected Variables</td>
<td>111</td>
</tr>
<tr>
<td>Table 5-2(c)</td>
<td>Mean Values of Key Variables by District</td>
<td>111</td>
</tr>
<tr>
<td>Table 5-3</td>
<td>Single Equation PSI Model Parameters:MODEL0</td>
<td>113</td>
</tr>
<tr>
<td>Table 5-4(a)</td>
<td>Deterioration Model Parameters:MODEL1</td>
<td>119</td>
</tr>
<tr>
<td>Table 5-4(b)</td>
<td>Maintenance Model Parameters:MODEL1</td>
<td>119</td>
</tr>
<tr>
<td>Table 5-5(a)</td>
<td>Deterioration Model Parameters:MODEL2</td>
<td>123</td>
</tr>
<tr>
<td>Table 5-5(b)</td>
<td>Maintenance Model Parameters:MODEL2</td>
<td>124</td>
</tr>
<tr>
<td>Table 5-5(c)</td>
<td>Measurement Model Parameters:MODEL2</td>
<td>124</td>
</tr>
<tr>
<td>Table 5-5(d)</td>
<td>Comparison of PSI and Extracted Latent Variable</td>
<td>124</td>
</tr>
<tr>
<td>Table 5-6(a)</td>
<td>Deterioration Model Parameters:MODEL2A</td>
<td>128</td>
</tr>
<tr>
<td>Table 5-6(b)</td>
<td>Maintenance Model Parameters:MODEL2A</td>
<td>129</td>
</tr>
<tr>
<td>Table 5-6(c)</td>
<td>Measurement Model Parameters:MODEL2A</td>
<td>129</td>
</tr>
<tr>
<td>Table 5-7(a)</td>
<td>Deterioration Model Parameters:MODEL3</td>
<td>133</td>
</tr>
<tr>
<td>Table 5-7(b)</td>
<td>Measurement Model Parameters:MODEL3</td>
<td>133</td>
</tr>
<tr>
<td>Table 5-8(a)</td>
<td>Deterioration Model 1 Parameters:MODEL4</td>
<td>136</td>
</tr>
<tr>
<td>Table No.</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Table 5-8(b)</td>
<td>Deterioration Model 2 Parameters: MODEL4</td>
<td>137</td>
</tr>
<tr>
<td>Table 5-8(c)</td>
<td>Maintenance Model Parameters: MODEL4</td>
<td>137</td>
</tr>
<tr>
<td>Table 5-8(d)</td>
<td>Measurement Model Parameters: MODEL4</td>
<td>137</td>
</tr>
<tr>
<td>Table 5-9(a)</td>
<td>Statistics of Extracted Performance Variables</td>
<td>141</td>
</tr>
<tr>
<td>Table 5-9(b)</td>
<td>Condition Prediction by MODEL2 and PSI</td>
<td>141</td>
</tr>
<tr>
<td>Table 5-9(c)</td>
<td>Condition Prediction by MODEL3 and PSI</td>
<td>141</td>
</tr>
<tr>
<td>Table 6-1(a)</td>
<td>Deterioration Model Parameters: MODEL2X</td>
<td>150</td>
</tr>
<tr>
<td>Table 6-1(b)</td>
<td>Maintenance Model Parameters: MODEL2X</td>
<td>151</td>
</tr>
<tr>
<td>Table 6-1(c)</td>
<td>Measurement Model Parameters: MODEL2X</td>
<td>151</td>
</tr>
<tr>
<td>Table 6-2(a)</td>
<td>Comparison of PSI Model and MODEL2X</td>
<td>153</td>
</tr>
<tr>
<td>Table 6-2(b)</td>
<td>Comparison of PSI Model and MODEL2</td>
<td>153</td>
</tr>
<tr>
<td>Table 6-3</td>
<td>Comparison of MODEL2X and MODEL4</td>
<td>155</td>
</tr>
<tr>
<td>Table 6-4(a)</td>
<td>Sensitivity of Condition to 1-yr increase in age</td>
<td>159</td>
</tr>
<tr>
<td>Table 6-4(b)</td>
<td>Sensitivity of condition to 1% increase in ADT</td>
<td>159</td>
</tr>
<tr>
<td>Table 7-1</td>
<td>Rating Scheme for Bridge Deck Evaluation</td>
<td>170</td>
</tr>
<tr>
<td>Figure No.</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Figure 1-1</td>
<td>The Infrastructure Management Process</td>
<td>15</td>
</tr>
<tr>
<td>Figure 1-2</td>
<td>The Latent Variable Model for Infrastructure Facility Deterioration</td>
<td>22</td>
</tr>
<tr>
<td>Figure 2-1</td>
<td>Graph for the Calculation of Deduct Points</td>
<td>40</td>
</tr>
<tr>
<td>Figure 2-2</td>
<td>List of Models from Infrastructure Literature Reviewed in Thesis</td>
<td>42</td>
</tr>
<tr>
<td>Figure 3-1</td>
<td>The Latent Variable Model Specification with Performance Maintenance Interactions</td>
<td>62</td>
</tr>
<tr>
<td>Figure 3-2</td>
<td>Specification of a MIMIC Model</td>
<td>64</td>
</tr>
<tr>
<td>Figure 3-3</td>
<td>Simultaneity of Deterioration and Maintenance</td>
<td>71</td>
</tr>
<tr>
<td>Figure 6-1</td>
<td>(-PSI) Plotted against ADT</td>
<td>160</td>
</tr>
<tr>
<td>Figure 6-2</td>
<td>Deterioration Prediction by Different Models: Endogenous Maintenance</td>
<td>162</td>
</tr>
<tr>
<td>Figure 6-3</td>
<td>Deterioration Prediction by Different Models: Exogenous Maintenance</td>
<td>164</td>
</tr>
<tr>
<td>Figure 7-1</td>
<td>Correspondence between Condition and States</td>
<td>172</td>
</tr>
<tr>
<td>Figure 7-2</td>
<td>Calculation of Transition Probabilities</td>
<td>176</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

The condition of highways in the United States has been attracting considerable attention over the past few years. The National Council on Public Works Improvement stated in a report to the President and Congress that there is "convincing evidence that the quality of America's infrastructure is barely adequate to fulfill current requirements, and insufficient to meet the demands of future economic growth and development" [National Council on Public Works Improvements, 1988].

The amount of money expended on the construction and maintenance of infrastructure facilities in the United States has declined steadily over the past thirty years. A report written by the "Rebuild America Coalition", which is a coalition composed of private and public sector organizations interested in improving the condition of the nation's infrastructure states that spending on infrastructure and other public works has decreased from 1.9% of GNP in 1953-1959 to 0.4% of GNP between 1980-1984 [Rebuild America Coalition, 1988]. Of the total expenditure on public works improvements, a larger and larger proportion is being spent on maintenance rather than on new construction, with the proportion rising from about 33% in 1960 to over 50% in 1985 [National Council on Public Works Improvements, 1988]. The critical issue facing highway agencies today is how to allocate the limited resources that are available for maintenance and rehabilitation so as to obtain the best returns for their expenditure.

The result of this situation is that there is now a need for methods to analyze the performance of infrastructure facilities. This performance analysis involves the collection of data to monitor the condition and usage of existing facilities, the development of models to predict the future condition of facilities and the identification of maintenance strategies over time for the facilities.
The framework for such an analysis and the components of this framework are described in detail in subsequent sections. This thesis focuses on one of the components of the framework described above, i.e. the prediction of infrastructure facility condition over time. The methodology developed in thesis is applied in the context of highway pavements; however, the techniques are directly applicable to other infrastructure facilities as well. Examples include bridges, rails, buildings and any other facility whose deterioration cannot be directly measured.

1.1 A Framework for Infrastructure Performance Analysis

The problem of allocating resources for maintenance and rehabilitation of infrastructure facilities has both temporal and spatial dimensions. Typically, an operating agency such as a state highway department has under its purview a network of facilities, and the funds available to maintain these facilities over a given year are limited. Also, the effects of maintenance in any one year are felt several years into the future; therefore the selection of a maintenance strategy cannot be static, but has to be evaluated over time. Specifically, the problem of selecting the most effective maintenance and rehabilitation strategy for a network of infrastructure facilities over some time period of interest has to address the following questions:

1) Which facilities need to be maintained?
2) What activity, or group or activities need to be performed on each facility selected in 1), and "how much" of each activity needs to be performed? The "how much" is an intensity measure, since the same activity can be performed to different extents. For example overlay of a pavement section is a single activity, but different thicknesses of overlay can be applied.
3) At what point in the time period of interest should these activities be performed?

The decisions entailed by the answers to the above questions are clearly interrelated. The
selection of facilities determines the choice of activities and the time of performance of these activities; the reverse is also true. The solution to the above problem therefore consists of choosing, from a very large number of combinations of facilities, activities and schedules, a subset that meets the constraints specified by the availability of resources and optimizes, under some definition, the "performance" of the network over the time period of interest.

In its most general form, the above problem, which can be called the strategy selection problem, is very complicated. Over the past ten years several researchers have developed models that solve different simplified versions of the general problem. Tsunokawa[1986] provides a review of strategy selection models. A more recent review has been compiled by Madanat[1989]. Though the specific assumptions, applicability and solution techniques vary widely between the different models, the general methodology employed by these models can be summarized as follows:

1) For each facility, define a set of possible maintenance strategies, which can be characterized by separate activities, different extents of the same activity (for example pavement overlay of different thicknesses) or simply by annual maintenance expenditures.

2) Predict the deterioration of the facility under each maintenance strategy, given some exogenously specified traffic and environmental characteristics.

3) Determine the costs incurred by the agency as a result of the application of each strategy and the costs incurred by the users due to the consequent pattern of deterioration (this concept of user costs is defined in very different ways by different models - in its simplest form it is a measure of the worst allowable condition of the facility).

4) For each facility choose the strategy that minimizes the total costs to the user and the agency, or optimize over the network subject to budget constraints.
The critical steps in the above methodology are (2) and (3), the prediction of future condition and the associated costs. The accuracy of an optimal strategy selection model depends upon the ability to (a) define a measure that adequately represents the "condition" or the "performance" of the facility; (b) develop a model that can accurately predict the change in this condition measure over time as a result of maintenance and exogenous variables that influence deterioration and (c) measure or model, as accurately as possible, the costs arising from maintenance and deterioration. Failure to properly implement these steps may result in incorrect choices of maintenance strategies, and consequently inefficient utilization of limited resources. Thus, the maintenance decisions made by an agency for a particular facility are driven by the underlying mechanistic processes that govern the deterioration of the facility, and the deterioration prediction model is a mathematical formulation of these processes.

The selection of an optimal maintenance strategy cannot therefore be decided in isolation, but has to be treated as a component of a general framework that includes data collection and deterioration prediction as other components. A comprehensive model system has to be able to include interactions between the different components of this framework. This framework for infrastructure performance analysis can be envisioned as shown in Figure 1-1 with three components: Data Collection, Impact Prediction and Strategy Selection [McNeil et al., 1989]. Each component is described below.

The data collection component deals with the gathering of data that is required for the impact prediction stage. Data is collected on the extent of facility damage (such as area or length of damage), on the causal variables that affect deterioration such as traffic, age, environmental variables and facility surface type, on the costs of performing maintenance, and the impacts on the user arising from deterioration of the facility. The data may be collected by visual
FIGURE 1-1: THE INFRASTRUCTURE MANAGEMENT PROCESS
inspection, through manual measurements or through automated techniques. The data collected, after suitable processing can be used as input to the impact prediction models. The data can also be used for monitoring purposes to validate model predictions after a model system has been implemented and has functioned for a while.

The impacts prediction component uses the information obtained during data collection to build models that predict the deterioration of the facility and the impacts to the users of the facility as a result of this deterioration. A deterioration model links a measure of condition of the facility to a vector of explanatory variables. The condition measure in its simplest form is just the extent of damage; more complex indices that combine the extents of different damage types may also be used. Similarly, a user impacts prediction model measures the impact of deterioration to the user of the facility as a function of the condition of the facility. The most rudimentary models of this sort for highway pavements calculate the tire wear or the fuel consumed as a function of pavement roughness. More complicated models can be envisioned where some measure of "user comfort" can be defined as a function of pavement condition; however, such models are currently not available in the literature since such measures of comfort are difficult to define.

Finally the strategy selection component, as discussed before, involves the choice of an optimal maintenance activity that maximizes the performance of the facility over time subject to various resource constraints.

Two issues are important in the consideration of the framework for infrastructure performance analysis. The first issue is one of interactions between the different components of the framework of Figure 1-1. The data collected in the first component is used as input to the impacts prediction stage; therefore, the accuracy with which the data is collected affects the
accuracy of prediction. Also, the impact prediction models can themselves be used to determine what data needs to be collected. For example, many highway agencies collect data on a large number of different damage types, and the effect of some of these damage types may overlap. If it is found that these damage types do not contribute to any new information about the deterioration of the highway over time, then time and money need not be expended in collecting data on these activities in the future. Similarly, a deterioration prediction model can be used to compare between an expensive but accurate automated data collection technique and a visual inspection method that is less expensive but more prone to error. The decision on which method to use, or how the two methods should be combined can be made on the basis of how successful each method is in predicting deterioration. The interactions between impacts prediction and strategy selection have already been discussed. The accuracy of a strategy selection model depends upon the accuracy of the associated impact prediction models. A comprehensive model system for the selection of maintenance and rehabilitation activities should recognize these interactions and account for them in the model formulation.

The second issue of importance is uncertainty. Each component of the framework described above has uncertainties associated with it. For example, the uncertainties in data collection can arise due to errors in the measuring devices, and from the inability to directly measure some variables (some environmental factors, for example). Errors in impact predictions can arise due to improper definition of a performance measure, or an incomplete model specification. Errors in strategy selection can occur due to difficulties in measuring user and agency costs. As a result of these errors, the framework of infrastructure performance analysis is not deterministic. The deterioration of the facility over time cannot be predicted precisely, nor can the impacts to the user. Any realistic approach should explicitly include the stochastic nature of deterioration and the effects of errors in data collection.
In summary, the above discussion isolates the following two points:

a) Optimization models for selecting maintenance activities cannot be developed in isolation; they have to be developed in conjunction with, or subsequent to impact prediction models.

b) Impact prediction models should explicitly include the stochastic nature of deterioration.

The following section describes the proposed approach for the development of impact prediction models, specifically in the context of pavement deterioration.

1.2 The Latent Variable Approach to Deterioration Modeling

The fundamental problem with the modeling of pavement deterioration is that there are no obvious units in which to measure the performance of the pavement. Therefore, there is no obvious way to characterize the dependent variable of the deterioration equation. Traditionally, this problem has been avoided either by selecting a small number of damage measurements as proxies for performance and predicting the propagation of each measurement independently, or by using inspector ratings of the pavement (on an arbitrary scale) and regressing these ratings against the damage to obtain a performance index formula. The problem with these methods is that there is no rigorous manner of deciding which measurements to model, and the reliance on inspectors’ ratings makes the performance index formula somewhat arbitrary and difficult to transfer across locations. In this thesis, new statistical techniques for rigorously estimating models of infrastructure deterioration are developed. The major contribution of this research is the development of a flexible model framework within which many different models can be specified depending upon the particular facility and the nature of deterioration. The necessity to a priori choose a subset of damage measurements, or to use inspector ratings to calculate an index formula is eliminated by the simultaneous estimation of a model that uses the mea-
measurements both to calculate a performance index as well as to predict the propagation of this index over time. The deterioration model in this approach is a structural equation system in which some of the variables are latent.

The concept of latent variables was introduced in the social sciences to model characteristics that are not easily measurable or directly observable in the population. Examples of such variables in the social sciences are "intelligence" or "verbal ability" in psychology, and "ambition" or "racial prejudice" in sociology. Though these concepts are not directly measurable, it is possible to observe the effects of these variables on certain other measurable or observable variables. These effects can be measured, either through tests such as standardized intelligence tests or verbal ability tests, or merely from observations on the population. For example, a measure of ambition could be the number of promotions or changes to better jobs in a certain number of years. These observable variables are therefore, in some sense, indicators of the underlying fundamental variable. Multiple indicators are usually available for a latent variable; for example, results of several intelligence tests can be used as measurements for the same underlying variable, i.e. intelligence.

The idea that there are some unobservable fundamental variables that describe observed phenomena is directly applicable to the modeling of highway pavement deterioration. One concept of latent phenomena in this context is the visualization of deterioration as the result of a variety of unobservable micro-mechanistic processes. A highway pavement may suffer damage due to many different processes: load related fatigue, shrinkage due to low temperatures, non-load associated distortion resulting from foundation movement and frost heave, disintegration of material and any interaction of the above [Haas and Hudson, 1978]. The "condition" of the pavement is affected by these processes, but since these are unobservable, pavement condition is unobservable as well. However, the effects of these processes are
observable as visible surface damage whose extents can be measured. These measured extents of damage then serve as indicators for the latent condition. Another viewpoint is to think of damage as the observed effect of some underlying unobserved characteristic that is closely related to the performance of the pavement. Under this viewpoint, the performance relates to the ability of the pavement to carry out the purpose for which it has been designed. For example, "load bearing capacity" or "safety level" or "ride quality" are three examples of pavement performance characteristics. Deterioration of the pavement corresponds to a degradation of it's performance, and this degradation can be observed by increased levels of surface damage. The different types of damage that are typically observed on a pavement (such as cracking, patching, rutting and potholes) are multiple indicators for unobserved performance.

The latent variable model of pavement deterioration specifies the deterioration model as a degradation of the latent performance variable over time; in other words, the deterioration model links the latent variable directly to the explanatory variables. Simultaneous to the deterioration model, a measurement model is specified that connects the observed variables with their latent counterparts. The general latent variable problem then involves the estimation of the latent variables from the indicators, using both the deterioration model and the measurement model.

The simplest latent variable model in the context of highway pavements (and, more generally, in the context of infrastructure facilities) can be specified as a natural extension to the deterioration models found in the literature. Recall that a traditional deterioration models links some index of condition to a vector of explanatory variables. A similar relationship can be specified for the latent variable model, except that the condition index, which is the dependent variable of the deterioration relationship, is unobserved. Because of the unobserved dependent variable, this relationship cannot be estimated by itself, but forms only one part of the latent variable
model. This part is denoted as the structural model. The second part of the latent variable model is the set of equations that are needed to define the relationship between the underlying condition of the pavement and the measurements of the indicators of this condition, i.e. the amount of damage on the pavement surface. This set of equations is called the measurement model. The latent variable method simultaneously estimates the parameters of the structural and the measurement model, thereby developing a deterioration model that is directly tied to the observed data. Figure (1-2) is a diagram of the major elements of the latent variable model.

More complicated models than the one described above can also be formulated. For example, the description in the preceding paragraph implicitly assumed that the condition of the pavement could be completely defined by a single latent variable. This need not necessarily be the case. Two dimensions of performance, characterized by two latent variables are easily possible; one relating to the sub-surface condition that determines the pavement's ability to bear loads (the "load bearing capacity") and the second related to the smoothness of the ride surface (the "ride quality"). Surface damage is the indicator for both these latent variables, though the actual damage types that serve as indicators for each latent variable may be different. For example, alligator cracking and rutting are the indicators for the load bearing capacity, while the slope variance or raveling are indicators for roughness. Clearly, some damage types might overlap, being indicators for both latent variables. A structural model can be specified with one latent variable being the function of another latent variable; for example, the ride quality could be a function of the structural strength of the pavement.

More complex specifications can also be introduced through the explanatory variables. Not all explanatory variables are exogenous; from the discussion of the previous chapter on the simultaneity of deterioration and maintenance, maintenance is clearly an endogenous explanatory variable, since the amount of maintenance performed is affected by the condition
Figure 1-2: The Latent Variable Model for Infrastructure Facility Deterioration
of the pavement. In addition, the explanatory variables can also be a mixture of latent and observed variables. A latent explanatory variable could be "environmental effects", which are not observed, and precipitation, minimum temperature or the number of freeze-thaw cycles can be used as indicators for the environmental effects. A set of measurement equations then need to be specified linking the latent explanatory variables to their indicators as well. Finally, even if an explanatory variable is observed, it could still be measured with error; the observed value can then be represented as the true value plus some random error term in a traditional econometric errors-in-variables specification [Kmenta, 1980]. Different specifications representing special cases of the general latent variable model are described in Chapter 3.

Uncertainty is introduced into the model through the inclusion of random error terms. There is one random error term associated with each measurement equation, and one associated with each causal equation. The error associated with the structural equations account for random fluctuations and unobserved effects in the deterioration process. The error terms in the measurement equations are a reflection of (1) specification errors associated with each indicator, since an indicator is only an imperfect proxy for the associated latent variable and (2) measurement errors reflecting errors in the measuring devices themselves.

1.3 Simultaneity of Deterioration and Maintenance

The second major contribution of this research is in the modeling of the interactions between deterioration and maintenance, which arise because maintenance is not exogenously specified but is performed in response to pavement condition. This interaction has not been considered in the deterioration models in the literature, leading to incorrect specifications and sometimes estimation of the wrong model.
The condition of a pavement at any time is the result of all the events that have taken place on the pavement in the past. If a controlled experiment is performed on a test section where the pavement condition is monitored by running different traffic loads on pavements maintained to different levels, then the condition at some time \( t \) can be exactly specified in terms of the events of the past, and the future deterioration of the pavement can be predicted from a time series model. The maintenance specification in this experiment is \textit{exogenous}, i.e. the activities and extents of maintenance to be performed are \textit{pre-specified} and do not depend on the condition of the pavement.

However, most of the data used for the estimation of deterioration models is collected on in-service pavements. For such pavements, maintenance activities are not specified exogenously, but are performed \textit{in response} to the condition of the pavement as a result of previous condition assessments. Thus, historical maintenance activities must be treated as endogenous variables, and every data point obtained from in-service pavements represents the outcome of two processes taking place simultaneously over time. The first process is the process of deterioration as a result of traffic loads, environmental factors and past maintenance. The second process is the maintenance that is performed over the years by the highway agency which is a function of past deterioration and the factors that might cause future deterioration, for example traffic loads. These processes can be summarized by the following equations:

\[
S(t) = f(A(t-1), A(t-2), \ldots, T(t-1), T(t-2), \ldots, S(t-1), S(t-2), \ldots, X_1(t)) + \eta_1(t) \tag{1-1a}
\]

\[
A(t) = g(S(t), S(t-1), \ldots, T(t), T(t-1), \ldots, X_2(t)) + \eta_2(t) \tag{1-1b}
\]

where

\( S(t) \) is the condition at time \( t \)

\( A(t) \) is the amount of maintenance performed at time \( t \)
T(t) is the traffic at time t

X_1(t) is a vector of explanatory variables for the deterioration equation at time t.

X_2(t) is a vector of explanatory variables for the maintenance equation at time t.

The system of equations (1-1) is recursive since the condition at time t depends upon the extent of maintenance at time t-1 and the maintenance at time t depends upon the condition at time t. If there is temporal correlation between the error terms, then the system of equations (1-1) behaves like a simultaneous equation model. Estimation of a single equation model when a simultaneous equation model is appropriate leads to biased parameter estimates.

The stationary version of these equations, written for any time period, can be expressed as follows:

\[
S = f(A, T, X_1) + \eta_1 \quad (1-2a)
\]

\[
A = g(S, T, X_2) + \eta_2 \quad (1-2b)
\]

Even if the simultaneous equation model is properly specified, it is not possible to estimate both equations in the system (1-1) unless both equations are identified. Consider the system in equation (1-1) for the case where X_1 and X_2 are identical vectors. In order for the deterioration equation to be identified, at least one independent variable that is contained in the maintenance equation should be omitted from the deterioration equation. Examination of equations (1-1) reveals that the traffic at time t, T(t), is one such variable. The deterioration equation is therefore identified. However, there is no variable in the deterioration equation that does not appear in the maintenance equation. The maintenance equation is therefore not identified unless X_1 and X_2 are different. In other words, in order for the maintenance equation to be identified, X_1
should contain some variables that are not found in $X_2$. For the stationary version of equation (1-2), non of the equations are identified if $X_1$ and $X_2$ are identical. Some researchers have reported [e.g. Butler, Carmichael and Flanagan, 1985] that "deterioration" specifications involving maintenance effects produce unexpected signs for important explanatory variables. It is possible that this occurs not because the specification is wrong but because in actuality the maintenance model is being estimated rather than the deterioration model. This issue is discussed in more detail in Chapter 3. Such problems might also exist with other deterioration models in the literature that are estimated from data collected on in-service pavements.

1.4 Contributions of research

In summary, the discussions of the preceding sections indicate the following requirements for a deterioration model:

1) The method should be able to estimate unobserved performance directly from measured damage data without relying on inspector ratings, and be able to choose between different measurements on a rigorous statistical basis rather than through the use of judgements. The inspector ratings themselves can be used as additional information to supplement observed data.

2) The method should be flexible enough so that no constraints need to be placed on the type of measurement of the data required for estimation. This makes it possible to evaluate different measurement techniques and data collection strategies.

3) Efficiency of estimation can be improved by a procedure that simultaneously estimates a performance index and a deterioration model since such a procedure maximizes the likelihood
over all parameters to be estimated. This is in contrast to a sequential process that is typically used in which separate regressions are performed to calculate the performance index and the deterioration relationship.

4) The deterioration model should include the simultaneous interaction between deterioration and maintenance.

5) The modeling approach should be probabilistic and should be able to include the effects of errors in measuring devices and the effect of past deterioration patterns on the present.

The contribution of this thesis is the formulation and estimation of a flexible model system for predicting highway pavement deterioration that satisfies the requirements outlined above. The results of implementing the latent variable models on field data indicate that the proposed approach produces realistic deterioration prediction models with the parameters having the expected signs, and a good overall fit to data.

This thesis consists of a total of eight chapters. The following chapter is a literature review of the infrastructure deterioration models found in the literature and a critique of their shortcomings. In Chapter 3, a latent variable model is formulated for highway pavements. In Chapter 4, the statistical issues relating to model estimation, measurements of fit and testing are discussed. Chapter 5 presents a case study that applies the latent variable approach to highway pavement deterioration data. Chapter 6 discusses the practical applications of the latent variable technique, including how it can be used for data collection and as an input into maintenance planning models. Chapter 7 extends the latent variable approach to the case of discrete damage data, and the methodology is described in the context of bridge decks. Finally, Chapter 8 presents the conclusions that can be drawn from this research and some suggestions
for further extensions. It is expected that the methods presented in this thesis, and their applicability in many different situations will provide a flexible analytic tool for the development of more realistic models for the allocation of resources over an infrastructure network.
CHAPTER 2

PAVEMENT DETERIORATION MODELS - STATE OF THE ART

This chapter introduces the principal concepts associated with the modeling of highway pavement deterioration. Modeling methods found in the literature, and some of the problems inherent in these models are discussed. As mentioned in Chapter 1, a deterioration model predicts the change in the performance of the pavement over time as a result of the factors that affect the deterioration of the pavement. There are two requirements for the specification of such a model. They are:

a) Definition and measurement of pavement performance.
b) Measurement of the factors affecting deterioration.

2.1 Performance of a Highway Pavement

The dictionary definition of the performance of a system is its "ability to meet a requirement, usually with regard to effectiveness" [Webster, 1979]. In the case of a highway pavement as well, the performance of the pavement can be defined as its ability to meet the requirements for which it has been designed. Since a pavement's primary purpose is to carry traffic loads comfortably and safely, the performance of a pavement therefore has to be intimately linked to the user's perception of the "quality" of the pavement. This was recognized in the earliest studies of pavement performance, when the concept of "pavement serviceability" was defined [Carey and Irick, 1960]. The serviceability of the facility is defined as a measure used to evaluate the facility from the user's perception of its condition. Other sources in the literature have also identified the concept of performance in terms of the ability of the pavement to fulfill its design requirements. For example, Haas and Hudson [1978] define "pavement condition"
as involving four major components: (1) riding comfort (2) load carrying capacity (3) safety and (4) aesthetics. Similarly, Kulkarni et. al. [1980], in a report of a study conducted for the Arizona DOT, define as "functional criteria" the broad areas of concern that are relevant to the determinations of acceptability of pavement performance. Such criteria are defined to be safety, riding comfort and physical distress.

A specification of a performance prediction model uses performance as a dependent variable, and a vector of the factors that affect performance (or deterioration) as the independent variables. The factors that affect deterioration vary from region to region depending upon the soil conditions and environmental factors. For example, Hajek and Haas [1987] identify the following four most important reasons for pavement damage in Ontario, Canada: traffic loads, temperature changes, moisture effects and construction flaws due to materials and construction techniques. Information is usually available on the age, pavement type and traffic flow, though the traffic information might only be available as a discrete traffic level (i.e. high, medium or low). Environmental data such as average annual rainfall, number of freeze thaw cycles, average annual temperatures etc. are also usually available by geographic region. The amount of maintenance performed also influences the performance of the pavement. Maintenance history information is mostly unavailable, since before the advent of computerized pavement management systems, highway agencies did not bother to record this information. The lack of this information places a severe impediment on the ability to develop accurate deterioration prediction models, since future deterioration patterns are intimately connected to past maintenance activities. More information on maintenance history will be available in a few years, as more and more pavement management systems are implemented.

While data on the explanatory (independent) variables can be obtained in one fashion or another, the major problem with the specification of a performance prediction model is that
the performance of a highway pavement is a latent variable. In other words, there is no obvious manner in which the dependent variable of a deterioration equation can be measured. Even if an attempt were made to measure performance by conducting a survey of riders' perception of the quality of the pavement, there is no natural scale, or natural units on which to quantify the users' perceptions. A method therefore needs to be found by which the performance of a pavement can be quantified. Once this is accomplished, then deterioration models can be specified that link this quantitative measure to explanatory variables.

The "quantifiable unit" that is easily measurable and that is directly correlated with the performance of a pavement is obviously the amount of damage on the pavement. The more the damage on a pavement, the worse is its ability to perform. Many deterioration models in the literature, therefore, have used surface damage as an indicator for the underlying, unobserved performance. In the early years of pavement performance research, when the concepts of serviceability were first being developed, Carey and Irick [1960] made the following observations:

"1) Highway users' opinion as to how they have been served by highways is by-and-large subjective.
2) There are, however, characteristics of highways that can be measured objectively and that, when properly weighted and combined, are in fact related to users' subjective evaluation of the ability of the highway to serve them."

These observations imply that the performance of a pavement can be quantitatively measured as a function of surface damage measurements. This is the principle upon which the state-of-the-art of pavement deterioration modeling is based. In the following section, the types of damage data collected and a short discussion of the data collection techniques is presented.
2.2 Collection of Pavement Damage Data

Data is generally collected by different highway agencies on some or all of the following:

1) Surface damage
2) Roughness
3) Deflection
4) Skid resistance

Of these, the data that is most widely used in deterioration modeling is the data on surface damage. The reason for this is that surface damage data is the easiest to collect. Data collection is done largely through visual inspection, supplemented by some degree of physical measurements. Pavements are divided into sections and data is collected on each section. The assumption made is that each section is an identical homogenous stretch of pavement, and that the distresses are uniformly distributed across the section. For example, data collected for the development of the PAVER maintenance management system [Shahin and Kohn, 1981] were on sections of 2500 sq.ft. The data collected by ARE inc. from Nevada for their deterioration models [Butler, Carmichael and Flanagan, 1985] used pavement sections one mile long. Data is collected on both the extent and the severity of damage. The extent of damage is the area or the percentage area of the section that is damaged, while the severity measures how serious the damage is. The types of damage on which data is obtained are numerous, and vary from agency to agency. Table (2-1), taken from Zarniewski [1985], shows the different types of damage data collected for asphalt pavements. As is evident from this table, state agencies vary widely on the number and kinds of damage data collected. The definition of severity is also different for different damage types. For example, the following are the definitions of severity levels for alligator cracking in the PAVER system:
<table>
<thead>
<tr>
<th>Damage Types</th>
<th>AL</th>
<th>AK</th>
<th>AZ</th>
<th>AR</th>
<th>CA</th>
<th>CT</th>
<th>FL</th>
<th>GA</th>
<th>IL</th>
<th>IA</th>
<th>IN</th>
<th>KS</th>
<th>KY</th>
<th>LA</th>
<th>ME</th>
<th>MN</th>
<th>NM</th>
<th>NY</th>
<th>NC</th>
<th>ND</th>
<th>OH</th>
<th>OK</th>
<th>OR</th>
<th>PA</th>
<th>RI</th>
<th>TX</th>
<th>UT</th>
<th>VA</th>
<th>WA</th>
<th>WI</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asphalt Cracking</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Bleeding</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Block Cracking</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Bumps and Sags</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Cracking</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Compressions</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Depressions</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Drainage</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Deep Track Ravel</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Edge Cracking</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Failures</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Frost Heaves</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Joint Reflection Cracking</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lane/Shoulder Drop Off</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Longitudinal Cracking</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Map Cracking</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Multiple Cracks</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Patches</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Polished Aggregate</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Potholes</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Profile Distortion</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Purging</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Pushing</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Pushing</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Railroad Crossing</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Random Cracking</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Rutting</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Shaving</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Slippage Crackling</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Swell</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Transverse Cracking</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Unslotted Patching</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Weathering and Raveling</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Widening Drop Off</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Number of Variables Observed</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>10</td>
<td>14</td>
<td>5</td>
<td>6</td>
<td>9</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>1</td>
<td>8</td>
<td>9</td>
<td>11</td>
<td>4</td>
<td>1</td>
<td>15</td>
<td>12</td>
<td>8</td>
<td>6</td>
<td>15</td>
<td>11</td>
<td>180</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2-1: Damage Types for Asphalt Pavements
**Low Severity**: Fine longitudinal hairline cracks running parallel to each other with none or only a few interconnecting cracks

**Medium Severity**: Further development of light alligator cracks into a pattern or network of cracks that may be lightly spalled

**High Severity**: Network or pattern cracking has progressed so that the pieces are well defined and spalled at the edges. Some of the pieces may rock under traffic.

With visual inspection, pavement damage data is largely based on the assessments of raters. Field raters inspect a pavement and based on their individual observations, record the damage data using evaluation forms. The specific documentation of the data varies from organization to organization. In New South Wales, Australia, for example, data is not collected on the individual damage types. Instead, the lateral and longitudinal extents of general damage is rated. On the other hand, Washington State DOT, the Texas State Department of Highways and Public Transportation, the Ontario Ministry of Transportation and the California Department of Transportation use inspection procedures that specifically document severity and extent of damage [Balta, 1984]. Another example of visual assessment is in Arizona, where pavement raters inspecting the percentage of cracking in a flexible pavement use a guide consisting of a range of photographs that denote various percentages of cracking. The rater then determines percent cracking by visually matching the pavement’s condition with one of the photographs in this standard guide. Sometimes some actual measurements are taken to supplement the subjective rating. In the PAVER system, for example, the inspector is provided with a hand odometer to measure distress lengths and areas, a 10-ft. straight edge, and a ruler to measure the depths of ruts and depressions [Shahin and Kohn, 1981].

Non surface damage measurements, such as those collected on pavement roughness, deflection or skid resistance are measured using mechanical instruments. There are many devices for
measuring ride quality; Maysmeter or a PCA roadmeters are the most common. The Maysmeter measures rear axle to body excursions through a photocell sensing system with a 0.1 inch resolution. This device drives a stepping motor for pen-and-chart drive movements of a 6-in. wide paper tape recorder. The recording pen moves at a rate proportional to the movements of the vehicle body and its differential. Distance is also indicated on the chart by an automatic event marker connected to the speedometer drive. The roughness measurement, which is directly proportional to the total body-differential movement, can be obtained by measuring the amount of chart movement per unit of road length traveled [Haas and Hudson, 1978]. Most roughness measurements are calculated in this fashion, as a function of the difference in height between the center-line and the wheel paths.

For measurement of the deflection of the pavement under a static load such as a vehicle with a loaded wheel, a Benkelman beam deflectometer or a falling weight deflectometer is used. The basic operation of these devices is similar. The tip of an 8 to 10 feet long beam probe is placed on the surface of the pavement very near the loaded wheel or in a position so that the loaded wheel will pass very close to it. As the loaded wheel moves by the probe tip, the movement of the pavement’s surface is recorded by the angular rotation of the beam probe.

For skid resistance, most agencies use either a Mu meter or an ASTM skid trailer. A skid trailer measures the force required to drag a non-rotating tire over a wet pavement. A Mu meter measures skid resistance in the "yaw mode", where the wheels are turned at some angle to the direction of motion. The side or cornering force is measured for wheels yawed at different angles.

Further details on measurement devices for pavement characteristics are provided in Haas and Hudson [1978].
2.3 Modeling Pavement Deterioration from Damage Data

As discussed in section 2.1, the data that is collected on surface damage are used as substitutes for unobserved performance in the specification of deterioration models. The most direct method by which this can be done is to use each individual damage measurement as a separate independent measure of performance. Under this approach, therefore, a separate model is estimated for the progression of each damage type as a function of the explanatory variables. For example, such a model system has a model that predicts the propagation of alligator cracking, another model that predicts the propagation of potholes and so on. As Table (2-1) shows, some agencies collect data on a large number of damage types, leading therefore to a large number of individual models. It is difficult to compare pavements on such a large number of dimensions when the deterioration models are used for the purpose of selecting maintenance activities on a pavement. Typically, highway agencies select from these models three or four "important" damage types and the models that predict the propagation of these damage types are used to plan maintenance activities [see, for example, Kulkarni et al, 1980]. The important damage types are based on judgement and the specific characteristics of deterioration in each state.

The second method for using surface damage data as substitutes for performance involves the calculation of a performance index from the observed damage data. The procedure by which this is done typically consists of the following two steps, performed sequentially:

STEP 1: inspector Ratings of Pavements: In this step, a research study is conducted, wherein engineers select a small representative set of damage measurements. This is then followed by a collection of a sample of inspectors' subjective ratings of the condition and measurements of the extents of the selected damage types. The ratings are then regressed on the measured
damage to obtain a performance index formula.

STEP 2: Model specification: In this step, data is collected on the different types of pavement damage for a large sample of pavements. For each pavement, the performance index is calculated using the formula developed in Step 1. The deterioration model is then estimated by regressing the performance index on explanatory variables.

In this method, termed the aggregate modeling approach, the connection between the users’ perceptions of the pavement and the observed damage is made in step 1, where it is assumed that the highway engineers have the experience required to distinguish between the performance of different pavements. Once the connection has been established between the inspectors’ evaluation and the damage, the model can be used simply by collecting damage data. The two most commonly used aggregate indices calculated by this method are the Pavilion Serviceability Index (PSI) and the Pavement Condition Index (PCI).

The PSI was developed as a result of the AASHO road test in 1961 [Highway Research Board, 1962] as part of a study to tie serviceability, i.e. the users’ perspective of the quality of the pavement to measurable surface damage factors. A group of evaluators were asked to provide their opinions on the quality of various ride surfaces and rate their opinion on an integer scale from 0 to 5, with 5 representing a newly resurfaced or constructed pavement and 0 representing a completely disintegrated surface. The terminology used for this rating was the Present Serviceability Rating (PSR), and the average of the evaluations by each rater was taken to be the PSR of a particular pavement. This rating was intended to simulate users’ points of view and was performed according to some ground rules discussed in Carey and Irick [1960]. The PSR was then regressed against damage variables transformed in different ways and the fitted
value of the PSR for each pavement was termed the Present Serviceability Index, or the PSI. Obviously, the PSI is also on a scale from 0 to 5. The estimated equation linking the PSI to observed damage for flexible pavements is as follows [Highway Research Board, 1962]:

\[
\text{Serviceability} = 5.03 - 1.91 \log_{10}(1 + \overline{SV})^{0.5} - 0.01(C + P)^{0.5} - 1.38(RD)^2
\] (2-1)

where \( \overline{SV} \) is the mean slope variance in utis of square inches obtained from a profilometer used for the road test and is a measure of roughness, C and P are the extents of cracking and patching in sq.ft. and RD is the mean rut depth in inches.

The second aggregate index of condition, the PCI was developed more recently by the U.S. Army Corps of Engineers for the PAVER maintenance management system [Shahin and Kohn, 1981]. The PCI is measured on a scale from 0 to 100, with 100 representing a new pavement. The PCI concentrates on surface damage components, and covers a larger number of damage types than the PSI. Nineteen damage types, with three severity levels for each, are used in the calculation of the PCI. The list of damage types is presented in Table (2-2). The extent of damage is converted into "deduct points" which is the number of points to be deducted from 100 to calculate the present condition of the pavement. These deduct points were estimated by experts who rated the condition of the pavements. Graphs such as those presented in Figure (2-1) are used to calculate the number of deduct points for simultaneously occurring multiple distress types. Many agencies use both indices, or some variant of these indices.

2.4 A Review of Pavement Deterioration Models

In the previous section, it was seen that the deterioration models that are found in the literature measure performance either by estimating a separate model for each damage type or by cal-
<table>
<thead>
<tr>
<th>Damage Code</th>
<th>Damage Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alligator Cracking</td>
</tr>
<tr>
<td>2</td>
<td>Bleeding</td>
</tr>
<tr>
<td>3</td>
<td>Block Cracking</td>
</tr>
<tr>
<td>4</td>
<td>Bumps and Sags</td>
</tr>
<tr>
<td>5</td>
<td>Corruption</td>
</tr>
<tr>
<td>6</td>
<td>Depression</td>
</tr>
<tr>
<td>7</td>
<td>Edge Cracking</td>
</tr>
<tr>
<td>8</td>
<td>Reflection Cracking</td>
</tr>
<tr>
<td>9</td>
<td>Lane/Shoulder Dropoff</td>
</tr>
<tr>
<td>10</td>
<td>Longitudinal/Transverse Cracking</td>
</tr>
<tr>
<td>11</td>
<td>Patching and Utility Cut Patching</td>
</tr>
<tr>
<td>12</td>
<td>Polished Aggregate</td>
</tr>
<tr>
<td>13</td>
<td>Potholes</td>
</tr>
<tr>
<td>14</td>
<td>Railroad Crossing</td>
</tr>
<tr>
<td>15</td>
<td>Rutting</td>
</tr>
<tr>
<td>16</td>
<td>Shoving</td>
</tr>
<tr>
<td>17</td>
<td>Slippage Cracking</td>
</tr>
<tr>
<td>18</td>
<td>Swell</td>
</tr>
<tr>
<td>19</td>
<td>Weathering and Raveling</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Severity Code</th>
<th>Severity Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Low</td>
</tr>
<tr>
<td>2</td>
<td>Medium</td>
</tr>
<tr>
<td>3</td>
<td>High</td>
</tr>
</tbody>
</table>

Table 2-2: Damage Types used for PCI Calculation
Alligator Cracking

Longitudinal Cracking

Figure 2-1: Graph for Calculation of Deduct Points

(from Shahin and Kohn, 1981)
calculating a performance index from the damage data, and then regressing this performance index on explanatory variables. In this section, some models of highway pavement deterioration in each of these categories that have been used or are being used by various agencies are described. A list of these models is presented in Fig (2-2).

A) SEPARATE MODELS FOR DIFFERENT DAMAGE TYPES:

ARE: The purpose of the ARE study [Butler, Carmichael and Flanagan, 1985] was to predict pavement damage and serviceability as a function of the pavement design, material characteristics, traffic, environmental factors and maintenance and rehabilitation treatments. Five damage types are modeled in the study for seven surface types and three base types. A total of forty separate damage models are used for prediction.

The five damage types modeled are:

a) Cracking
b) Raveling
c) Potholes
d) Rut depths
e) Roughness

The factors affecting deterioration are a mixture of exogenous variables, and pavement related factors. The data requirements for the model can be categorized into the following groups:

1) Pavement characteristics: Pavement strength, layer thicknesses, age, base type and surface type
A) SEPARATE MODELS FOR DIFFERENT DAMAGE TYPES:

1) ARE
2) EAROMAR
3) Roughness Prediction Models
4) Arizona DOT Model

B) MODELS BASED ON PERFORMANCE INDICES

1) PARS Model
2) Serviceability and Distress Model
3) Continuous Deterioration Model
4) PAVER
5) Carnahan

Fig 2-2: List of Models from Infrastructure Literature
Reviewed in Thesis
2) Pavement history: Time since last rehabilitation, total pavement age

3) Pavement condition: Extent and severity of cracking, rutting, potholes, roughness and raveling

4) Environment: Annual average monthly precipitation

The ARE deterioration models are adapted from the highway design and maintenance (HDM) model system developed by the World Bank [Paterson, 1987]. Two different sets of models are developed for different phases in the deterioration process. The first phase is the initiation phase, which is the period before surface damage of a given type or severity appears; the second phase is the progression phase, during which the extent of damage progresses over time. For each time period and damage type, these models predict the change in condition. The condition for each time period is then updated by adding this change in condition to the previous condition:

Maintenance and rehabilitation effects are introduced into the model through a "distress retardation factor" which is the amount by which the particular maintenance activity retards deterioration. This factor was estimated by judgement. The predicted change in condition without maintenance is multiplied by this factor to give the adjusted change in condition. This adjusted change is then added to the old condition in eqn (2-2) to give the new condition.

EAROMAR: The EAROMAR model system [Markow and Brademeyer, 1984] is a simulation model to predict performance and maintenance and rehabilitation costs. The EAROMAR approach is similar to the one used by ARE, except that deterioration is predicted by modeling the change in the sub-surface material properties as a result of traffic loads and precipitation rather than simply specifying condition as a function of a vector of explanatory variables. Extensive amounts of data are required as input to the simulation. The performance prediction component requires inputs on route characteristics, traffic demand, environmental conditions,
maintenance policies and pavement characteristics. The effect of maintenance on retarding deterioration is modeled in a much more direct fashion than in the ARE models. Rather than express the amount of maintenance performed in traditional units of output, such as square yards of activity, annual expenditure per year or labor and equipment resources consumed, maintenance input in the EAROMAR models is directly expressed in terms of correcting damage. As a result, the effect of maintenance on damage can be directly calculated without the use of retardation factors. For example, the maintenance policy can be expressed as follows: "Repair 50% of the pothole damage in year t". The adjusted distress in year t is then calculated as the predicted pothole damage minus the 50% that is repaired.

EAROMAR is the only model in the literature that treats maintenance in this fashion. The problem with this approach is that it is not possible to measure effectiveness of different maintenance efforts, since the returns obtained from a unit investment of resources is not a linear function of damage. The "repair 50%" criterion is a different policy depending on the amount of damage present on the pavement. For example, repairing 50% of the pothole damage might involve a greater investment when 80% of the pavement surface has potholes, as compared to when only 20% of the pavement surface has potholes. This is because a pavement with severe surface damage would respond less to a routine maintenance activity than a pavement with small or moderate surface damage. As a result, the returns in terms of damage correction and future damage prevention obtained for each dollar spent on pothole filling are different depending on the overall level of surface damage. The EAROMAR model treats these policies as identical, i.e. it assumes that the effectiveness of repairing 50% of the pothole damage is the same, irrespective of the amount of damage present.

**Roughness Prediction Models:** To this class belong the many different models that predict the propagation of roughness over time. These models have as the dependent variable some
measure of roughness such as the Maysmeter or profilometer readings or slope variance and as the independent variable some combination of age, traffic and extents of other damage types such as cracking, rut depth and patching. Models of this sort have been developed by AASHTO[1972], Lytton et. al.[1982], Way and Eisenberg[1980], Potter[1982] etc. and are described in detail in Paterson[1989]. Paterson[1989] also formulates a roughness propagation model that predicts the change in roughness, $\Delta R$ using the following function:

$$\Delta R = f_1(\text{strength, condition, \Delta traffic, \text{environment}}) + f_2(\Delta \text{surface condition, \Delta maintenance})$$

$$+ f_3(\text{condition, environment, \Delta r}) + \epsilon$$

(2 - 2)

The surface condition is measured by the extent of cracking in percentage area of pavement section and the present roughness in units of metres/km. The change in condition is measured by the change in cracking and the change in rut depth standard deviation which is an indication of the change in roughness. The maintenance activity is surface patching in units of percentage area of section. The functions $f_1$ and $f_3$ are products of the independent variables, and $f_2$ is a linear function. The model includes exogenous variables that are not usually included in models of this sort in the literature, such as the effects of other damage types, environment and maintenance. Different models were estimated for various traffic categories, pavement strengths and maintenance policies.

**Arizona DOT Model:** The models described previously make predictions based upon the expected value of condition; i.e., after estimation, the random error term is discarded. Unlike these, the Network Optimization System (NOS) developed by the Arizona State DOT [Golabi, Kulkarni and Way, 1982] makes probabilistic statements about future condition. This system has been used to formulate models to select optimal maintenance strategies for the Arizona highway system. In this model, the condition at any point in time is represented by a four
component state vector. The components that describe the state of the pavement at any point in time are: present roughness, present cracking, change in amount of cracking during the previous year, and the index to the first crack. Rutting is not included in the formulation because it is not a problem in Arizona. The "index to the first crack" serves to distinguish between pavements of different overlay thicknesses; pavements with a higher value of the index to the first crack have thicker overlays.

Each component of the four dimensional vector can take a number of discrete values; present roughness, present cracking and change in cracking take three values each, while the index to the first crack can take 5 values. This makes a total of \((5 \times 3 \times 3 \times 3 =)\) 135 possible states that the pavement can occupy. Of these, 15 states are infeasible (e.g. present cracking < 10% with >15% change in cracking the previous year), so a final total of 120 states is obtained. A deterioration of the pavement involves a transition from the present state to a more deteriorated one. Since the future state of the pavement given the present state is a probabilistic process, the pavement can go from the present state to a number of "more deteriorated" states with some associated finite probability. This probability associated with each possible transition; this probability is termed the transition probability.

The Arizona model assumes that the propagation of distress over time follows a Markov process; therefore, the probability of transition from a given state to a more deteriorated state depends only upon the present state variables and the changes in these variables. The only source of "memory" in the process is introduced through the index to the first crack; the value of this index determines how long back the pavement was rehabilitated or the extent of rehabilitation, and thereby its future deterioration characteristics. The index to the first crack is therefore used as an explanatory variable in determining the extent of future cracking. For example, the equation for calculating the change in the amount of cracking for a given year for
an overlaid pavement is as follows: [Kulkarni et. al., 1980]

\[ C_N = 0.507 + 0.0687C_o + 0.52C_P - 0.0034I_e^2 - 0.003C_o^2 + 0.0681C_P^2 + \varepsilon \]  \hspace{1cm} (2-3) 

where \( C_N \) is the change in cracking during the next year, \( C_P \) is the change in cracking during the present year, \( C_o \) is the present cracking, \( I_e \) is the index to the first crack and \( \varepsilon \) is a stochastic error term. The Markovian nature of the process is apparent from this equation since future deterioration in any year predicted by this equation depends only upon present damage and the change in damage during the current year, given a specific value of the index to the first crack.

When the model is used for prediction, future condition is measured by the probability of transition from the present state to a lower state. If it is assumed in equation (2-3) above that \( \varepsilon \) is normally distributed, then the predicted future change in cracking, \( C_N \) is also normally distributed with a mean given by the deterministic part of equation (2-3) and a variance that is consistently estimated by the standard error of the regression. By discretizing the normal distribution of \( C_N \), the probability that \( C_N \) is within a particular range is calculated. Since the states are pre-specified to lie between certain ranges of cracking the above probability is the transition probability of deteriorating to another state. An equation similar to equation (2-4) is formulated for roughness as well. Deterioration due to roughness is assumed independent of deterioration due to cracking. The joint transition probability is therefore the product of the transition probability due to cracking and that due to roughness. Further details about the model are provided in Kulkarni et. al.[1980].

B) MODELS BASED ON PERFORMANCE INDICES:

Rather than estimate an independent deterioration equation for each distress type, the following
models predict the progress of a performance index over time.

**PARS Model:** The PARS [program and financial planning in pavement rehabilitation] model was developed for the state of Ontario [Kher and Cook, 1985]. This model uses PCI as the aggregate measure of condition, and predicts pavement condition as a function of pavement age, traffic and thickness of overlay. The model has the following form:

\[
PCI = 95 - k(\text{age})^\alpha (\text{thick})^\beta (\text{ADT})^\gamma
\]  

(2-4)

where age is the time in years since last rehabilitation, thick is the thickness of overlay, ADT is the average daily traffic, and \(\alpha, \beta, \gamma, k\) are parameters. The PCI is calculated from visual distress measurements as in the PAVER system.

**Serviceability and Distress Models:** These models were developed by [Garcia-Diaz and Riggins, 1984]. A set of models with identical functional forms are specified for predicting the progress of both PSI and damage:

\[
\text{PSI} = \text{PSI}_o - (\text{PSI}_o - \text{PSI}_f) \exp[-(\rho/W)\beta] \\
A = A_o - (A_o - A_f) \exp[-(\rho/W)\beta]
\]  

(2-5)  

(2-6)

where \(A\) represents area damaged for each of seven damage types. The damage types modeled are rutting, flashing, raveling, alligator cracking, transverse cracking, longitudinal cracking and patching. \(A_o\) and \(\text{PSI}_o\) represent the initial PSI and the initial damage extents for each damage type, \(A_f, \text{PSI}_f\) are the final condition, and damage extents, \(\rho, \beta\) are parameters, and \(W\)
is the cumulative traffic load measured as some function of the number of months the pavement has been in service since construction. Seven types of distresses are modeled. They are: rutting, flushing, raveling, alligator cracking, transverse cracking, longitudinal cracking and patching. The functional form gives a S-shaped curve which some researchers have shown is a good model for damage progression, reflecting a slow rate of wear for a period when the pavement is new, an escalation of damage in mid-life, and a slowing down again when the pavement is substantially deteriorated.

**Continuous Deterioration Model:** Balta[1984] and Fernandez[1979] used a model that predicts the deterioration of the pavement as a continuous function over time, as part of a strategy selection formulation that can be solved using optimal control theory. The model form is as follows:

\[
\frac{d\text{PSI}(t)}{dt} = \alpha \text{PSI}(t) + \beta \text{ESAL}(t) + \gamma A(t) \tag{2-7}
\]

where \( \frac{d\text{PSI}(t)}{dt} \) is the rate of change of (continuous) PSI at time \( t \), PSI\((t)\) is the PSI at time \( t \), ESAL\((t)\) is the traffic in ESAL and A\((t)\) is the annual maintenance expenditure in dollars per mile. The parameters \( \alpha, \beta, \gamma \) however were not estimated from data, but were estimated from other studies and judgement. Equation (2-7) is a differential equation for condition, and is the limiting case of the difference equations modeling change of condition described in Paterson’s roughness models or Arizona’s NOS system. Other models of this type have been specified by Tsunokawa[1986], and for rail deterioration by Murakami and Turnquist[1985].

**PAVER:** The PAVER system [Shahin and Kohn, 1981] uses the following predictive equation in terms of PCI:
\[ PCI = 100 - AGE \left[ 1.487/(\alpha_{zz}) + 0.143AGECOL + 6.56/T_{ac} - 1.23\alpha_{sc} \right] \]  

(2-8)

where AGECOL is the age between the time the pavement was constructed and its last overlay, 

\( T_{ac} \) is the total asphalt concrete thickness in inches, including overlay and \( \alpha_{zz}, \alpha_{sc} \) are functions representing base and sub-base characteristics, the tire contact area and tire pressure of an equivalent single wheel. The literature does not explain how the parameters were estimated.

**Carnahan:** Carnahan[1987] and others developed a stochastic model of pavement deterioration which uses the PCI as an aggregate index of performance. Like the Arizona DOT model, deterioration is represented as a Markov process with transitions taking place from one discrete condition state to another. The discrete states of the facility are defined by dividing the continuous PCI scale into discrete intervals. The transition probabilities are estimated from the numbers of observed transitions for different combinations of traffic and environment and for different rates of deterioration within each combination. A different model is therefore estimated for each exogenous variable combination. This categorization might make the model insensitive to small variations in explanatory variables.

The review above classified models into two groups: models where a separate equation was estimated for each damage type and models that predicted deterioration using a performance index. From the above description, several other possible categorizations for the models in the literature becomes apparent. They are:

a) theoretical or empirical

b) static or dynamic

c) expected-value or probabilistic
A theoretical model is one that explains the initiation and propagation of deterioration by modeling the effects of traffic, environment and climatic factors on the mechanical properties of the material. Examples are models that predict the structural response of a multi-layered pavement to traffic loads, or the change in elastic moduli of the pavement as a result of temperature changes. Theoretical models try to represent the physical processes that cause deterioration and predict future deterioration by the progression of these processes. The problem with models of this sort is that it is difficult to correlate changes in sub-surface properties to observed surface deterioration. This is because the effects of precipitation or of temperature changes on the material properties are complicated and may vary from location to location depending upon the soil properties and environmental conditions. As a result, even though a theoretical model is ideally desirable, not many such models have been implemented. An example is the EAROMAR model. More common are empirical models that predict the propagation of surface damage by regressing the observed or measured damage on explanatory variables and maintenance actions. These models are not explicitly based on any theory of material degradation, but correlate performance or condition to the factors that affect deterioration such as age, traffic or inches of precipitation. The ARE models or the Arizona DOT model system belong to this category.

Models can also be classified as static or dynamic. A static model simply predicts the condition of the pavement at various points in time as a function of explanatory variables. Such an approach plots a curve of best fit through different scattered observations of condition and the explanatory variables. The PAVER deterioration model, or the PARS model that specifies the PCI as a function of explanatory variables are examples of static models. A dynamic model predicts changes in condition that are added to previous condition to estimate the new condition. Typically, the change in condition is specified as being sensitive to both present condition and
the values of the exogenous variables at that condition. This is a representation of a Markov process. The Paterson roughness models or the continuous differential equation based models of Fernandez or Tsunokawa are examples of dynamic models.

Finally the models can be classified as expected-value based or probabilistic depending on the method used for predicting future condition. Expected-value based models ignore the probability distribution of future condition and simply predict its expected value. Probabilistic models, on the other hand, such as the Arizona DOT model or the models developed by Car- nahan, take into account the probabilistic nature of deterioration. In these models, the future condition can lie in one of many discrete states, with an associated probability of being in each of these states.

2.5 Limitations of deterioration models

The following are the limitations of the models presently used. The limitations are presented broadly classified by category of model described above:

1) Separate model for individual damage types: The data requirements for models such as the EAROMAR or the ARE models that separately predict the propagation of different damage types are very large. The ARE system contains over forty separate models, and for some of these models, parameters had to be guessed using judgement since data was not available to estimate them. The data requirements have also prevented the EAROMAR models from being widely used. In addition to the data limitations, the large number of damage models makes it difficult to use the system for maintenance planning since no method is suggested in the ARE or EAROMAR framework for selecting those damage types that are important for strategy selection.
2) **Models based on performance indices:** The aggregate indices used for quantifying pavement condition, such as the PSI or the PCI are based on judgements made by inspectors on the condition of the pavement. While experienced inspectors are capable of distinguishing between various levels of deterioration on a pavement, errors are still likely to arise in the course of the process of quantification, since the value of a condition rating to assign to a particular pavement could differ from inspector to inspector. Also, many of the existing indices of condition such as the PSI or the PCI are built up from a very specific set of damage components measured in very specific ways. For example, the PSI is calculated from the extents of cracking, patching, rut depth and slope variance. However, as Table (2-1) shows, highway agencies measure many different types of damage depending upon the geographic and environmental characteristics of the area, and these measurements may be different than those required for the PSI. Performance index based deterioration models therefore cannot be transferred easily. Also, a specific highway agency might want to incorporate other damage measurements in the calculation of a performance index. This is not possible under any of the current techniques, unless the agency chooses to define its own index by sending out a panel of raters to rate the pavements. This is time consuming and expensive, and the results obtained cannot be compared with existing indices.

3) **Inclusion of measurement errors:** None of the models described in this review explicitly include the errors that are present in the measurements of damage and the explanatory variables. In reality, substantial errors in measurement are present in data collected through visual inspection, especially in cases where the extent of damage is estimated while driving along the pavement. Even where detailed inspection of the pavement is done, area or length measurements are subject to errors because it is difficult to clearly demarcate where the damage begins and where it ends. Also, if the damaged area is irregularly shaped, it is difficult to precisely measure
the area of damage. If these errors is large, then predictions of condition based on models that
do not include these errors during estimation may be substantially different from the actual
conditions.

4) **Markovian assumptions of probabilistic models:** The models developed by the Arizona
DOT and Carnahan are based on the assumption of a Markovian deterioration process. In the
Arizona model, an allowance is made for differences in pavement construction quality through
the introduction of the "index-to-first-crack" variable. The assumption is that differences
between pavements exist only so long as cracking has not taken place; once cracking begins,
the behavior of every pavement under similar traffic, maintenance and environmental condi-
tions is identical. This might be an adequate assumption for the purposes of the Arizona DOT;
however, by tying the entire approach to the Markov assumption, the model does not have the
flexibility to deal with a more general non-Markovian deterioration process, of which the
Markov process is only a special case.

5) **Limitations of Static Models:** A static model such as the PARS model predicts a particular
value of condition for a given combination of explanatory variables. In a two-dimensional
framework, deterioration over time is represented by a single line. It is difficult to use this
model to predict future condition for points outside the line. Kher and Cook [1985] use the
slope of the regression line at the observed pavement condition to predict future deterioration
from the PARS model. This is equivalent to assuming that the slope of the deterioration equation
is the same for a given value of condition irrespective of the age of the pavement or traffic
loads. A dynamic model such as the ARE or the EAROMAR model is more realistic because
the change in condition is sensitive to both present condition and the value of the explanatory
variables at that condition.
A realistic model for deterioration prediction is therefore a probabilistic dynamic model that does not rely on subjective inspector ratings for the calculation of a performance index and which can be used to select the damage types that are important for maintenance planning. The latent variable approach presented in Chapter 1 provides such a framework. The formal specification of a latent variable model is presented in the following chapter. Because of data restrictions, only a static model is formulated and estimated in this thesis; however, changing the specification to a dynamic model that predicts changes in condition is possible without changing any of the major concepts of the approach.
CHAPTER 3
FORMULATION OF LATENT VARIABLE MODELS

At the end of Chapter 2, the requirements for a new approach for modeling infrastructure deterioration were listed. The main requirement was for a model framework within which unobserved performance could be directly estimated from observed damage measurements without relying on a performance index based on subjective ratings. The latent variable approach described in Chapter 1 provides such a framework. The formulation of a latent variable model for highway pavement deterioration is described in this chapter.

3.1 Specification of a General Latent Variable Model

The following description of the general latent variable model has been adapted from Everitt[1984] and Bartholomew[1987]. Let the vector \( Z \) consisting of elements \([z_1, z_2, \ldots, z_M]\) represent the observed variables and the vector \( L \) consisting of \([l_1, l_2, \ldots, l_p]\) represent the latent variables. The number of latent variables, \( P \) is typically much smaller than the number of manifest variables \( M \).

Let the joint multivariate probability distribution of \( Z \) given \( S \) be given by \( f_{Z|S}(z | s) \). If the observed variables are all continuous, then if the density of \( S \) is \( f_S(s) \), the unconditional density of \( Z \), \( f_Z(z) \) is given by

\[
f_Z(z) = \int f_{Z|S}(z | s)f_S(s)ds \tag{3-1}
\]

The information that is of interest from equation (3-1) is what can be known about \( S \) once \( Z \) is observed. This information is conveyed by the conditional density \( f_{S|Z}(s | z) \) which can be
calculated from $f_Z$, $f_S$ and $f_{Z|S}(z \mid s)$ using Bayes’ law. However, in order to find $f_{S|Z}$, it is necessary to know $f_{Z|S}$ and $f_S$, while all that is observable is $f_Z$. At this level of generality, therefore, it is not possible to obtain an estimate for $f_{S|Z}$. Also, the distributions $f_S$ and $f_{Z|S}$ of equation (3-1) that generate the distribution $f_Z$ need not be unique, since many combinations of $f_S$ and $f_{Z|S}$ gives rise to the observed distribution of $Z$. Therefore, the choices of $f_S$ and $f_{Z|S}$ are arbitrary, and can be chosen in any form that is convenient.

Some restrictions do apply on the choice of the prior distribution $f_S$ and the conditional distribution $f_{Z|S}$. The ability to represent $M$ manifest variables by a smaller set of $P$ latent variables arises from an assumption known as the axiom of conditional independence. Since the observed variables $Z$ are related to the latent variable $S$, the $Z$'s are correlated with each other. If the only reason for the correlations among the $Z$ values is their relation to the latent variables, then, if all the latent variables are held fixed, the $Z$ values are uncorrelated with each other. The object of the latent variable problem, therefore, is to find $P$ latent variables (where $P \leq M$) such that

$$f_{Z|S}(z \mid s) = \prod_{i=1}^{M} f_{Z_i|S}(z_i \mid S) \tag{3 - 1}$$

Bartholomew[1987] shows that an assumption of a normal distribution for $f_{Z|S}$ and $f_S$ produces a normal posterior distribution $f_{S|Z}$. In practice, it is assumed that the distributions of $f_{Z|S}$ and of $f_S$ are known, but are dependent on a set of unknown parameter values. In such a case, the problem of inferring $f_{Z|S}$ and $f_S$ from $f_Z$ becomes that of estimating the unknown parameters. Once the parameters are estimated, expressions for the distribution, or at least the first and second moments of the posterior distribution $f_{S|Z}$ can be calculated.

Specifically, suppose the prior distribution of $S$ is $P$-variate normal with mean zero and variance covariance matrix $\Psi$, and the conditional distribution $f_{Z|S}(z \mid s)$ is $M$-variate normal with mean
\( \lambda S \) where \( \lambda \) is a \((M \times P)\) matrix of parameters. From equation (3-1), the variance of the conditional distribution can be expressed by the diagonal matrix of variances \( \theta \). Then \( f_z \), the joint distribution of the observed variables is a M-variate normal distribution with mean zero and variance covariance matrix \( \lambda \Psi \lambda' + \theta \). The posterior conditional distribution \( f_{s \mid z}(s \mid z) \) is a P-variate normal distribution with mean \( \Psi \lambda' (\lambda \Psi \lambda' + \theta)^{-1} Z \) and variance covariance matrix given by \( (\Psi \lambda' (\lambda \Psi \lambda' + \theta)^{-1} \lambda \Psi) \) [Bartholomew, 1985]. The estimation problem in this case is to estimate the parameter matrices \( \lambda \), \( \Psi \) and \( \theta \) from observations on \( Z \).

The model just described above is termed the **linear latent variable model** since the regression of \( Z \) on \( S \) is linear in the expression for the mean of the posterior distribution. Though the development above assumed that all distributions above are multivariate normal, the assumption of normality is not critical for the derivation of the first and second moments of the conditional prior and posterior distributions. Consider any distribution \( f_z \) with mean zero and variance covariance matrix \( \Psi \) and any distribution \( f_{z \mid s}(z \mid s) \) with mean \( \lambda S \) and variance covariance matrix \( \theta \) that satisfy equation (3-1). The relation between \( S \) and \( Z \) can then be written as follows:

\[
Z = \lambda S + \epsilon \quad (3-2a)
\]

where \( E(\epsilon \epsilon') = \theta \), and \( E(\epsilon) = 0 \).

From equation (3-2a), \( E(Z) = E(S) = 0 \), and the variance of \( Z \) is equal to \( \lambda \Psi \lambda' + \theta \). \( f_z \) is therefore an M-variate distribution with mean zero and variance covariance matrix \( \lambda \Psi \lambda' + \theta \).

A standard result from econometric theory states that any vectors \( Z \) and \( S \) with a finite variance covariance matrix can be written in terms of a linear regression of \( S \) on \( Z \) as follows [Amemiya, 1985]:

58
and as long as $E(v)=0$ and $E(Zv)=0$, $E(S|Z) = \alpha Z$. $\alpha$ is then given by the usual least squares expression $\Sigma_{ZS}^{-1} \Sigma_{ZZ}$, where $\Sigma_{ZS}$ is the covariance matrix between $Z$ and $S$, and $\Sigma_{ZZ}$ is the variance covariance matrix of $Z$.

From equation (3-2a), $\Sigma_{ZS} = \Psi \lambda'$, and $\Sigma_{ZZ}$ is, as given above, $\lambda \Psi \lambda' + \Theta$. $E[S|Z]$ is therefore equal to $\Psi \lambda' (\lambda \Psi \lambda' + \Theta)^{-1} Z$, which is the same expression obtained in the case of the normal distribution. The latent variable model can therefore be estimated for distributional assumptions that are different from the normal. The exact nature of the posterior distribution is unknown in this case, but the first and second moments can be calculated as above.

Estimation of the parameters of the latent variable model is done by comparing the variance covariance matrix of the observed variables $Z$ calculated as a function of the parameters $(\lambda \psi \lambda' + \Theta$ with the observed (sample) variance covariance matrix of $Z$. The parameters are then estimated so as to minimize some measure of distance between the two variance covariance matrices. Estimation is described in greater detail in Chapter 4.

3.2 Formulation of the Latent Variable Model for Highway Pavements

Consider a highway pavement whose condition at some point in time can be represented by the latent (Px1) vector $S$. Suppose also that the vector of observed indicators is represented by the (Mx1) vector $I$. Finally, denote by the $K_1 \times 1$ vector $X_1$ the explanatory variables that affect deterioration, by $A$ the $K_2 \times 1$ vector of the extents of maintenance performed by different
activities and by \( X_2 \) the \( L \times 1 \) vector of explanatory variables on which the maintenance activities depend. The vector \( Z \) in the previous section is made up of elements of \( X_1, X_2, A, I \). The latent variable model for highway pavements can then be formulated as follows:

\[
S = g_1(A, X_1; \beta) + \eta_1 \tag{3-3a}
\]

\[
A = g_2(S, X_2; \gamma, \Gamma) + \eta_2 \tag{3-3b}
\]

\[
I = h(S; \lambda) + \epsilon \tag{3-3c}
\]

where

\( \beta, \lambda, \gamma, \Gamma \) are vectors of parameters, and \( \eta_1, \eta_2, \epsilon \) are random error terms.

Equations (3-3a) and (3-3b) specify the structural model that establishes the connection between the latent variables and the explanatory variables. In Chapter 1, the issue of simultaneity between condition and maintenance was discussed. Equations (3-3b) represent the maintenance equation that expresses the extent of maintenance performed measurement model expressing the indicators of damage as a function of the latent performance variables. Finally, equations (3-3c) form the measurement model, that links the observed damage with the latent variables.

Different assumptions about the functions \( g \) and \( h \) give rise to different models of pavement deterioration. The choice of functional forms would usually be made on the basis of observations of the deterioration of the pavement. In this case, however, the condition of the facility is latent and so no empirical evidence exists for the functional forms of \( g \) and \( h \). Since no models of this sort have been previously estimated in the literature, a linear model specification
is presented in this thesis as a first order approximation to more general functional forms. Linear functional forms are also convenient to estimate. Equations (3-3) can be written as follows for the linear latent variable deterioration model for highway pavements:

\[ S = \beta_1 A + \beta_2 X_1 + \eta_1 \]  
\[ A = \gamma S + \Gamma X_2 + \eta_2 \]  
\[ I = \lambda S + \epsilon \]  

(3-4a)  
(3-4b)  
(3-4c)

The estimation of the model specified by equation (3-4) involves the simultaneous estimation of the, the \((P \times K_1)\) matrix \(\beta_1\), the \((P \times K_2)\) matrix \(\beta_2\), the \((K_2 \times P)\) matrix \(\gamma\), the \((K_2 \times L)\) matrix \(\Gamma\) and the \((M \times P)\) matrix \(\lambda\). All the models estimated in the case study described in Chapter 5 are different forms of the basic specification of equations (3-4). The specification is depicted pictorially in Fig (3-1).

The model specified in equations (3-4) is complicated because of the endogenous explanatory variables in the structural equation system. The basic model from which this specification arises is called a **Multiple Indicator Multiple Causes** (MIMIC) model and is simply formulated by writing equations (3-4) without the maintenance model. The MIMIC model can be formulated as follows:

\[ S = \beta X + \eta_1 \]  
\[ I = \lambda S + \epsilon \]  

(3-5a)  
(3-5b)

where \(X\) is a \(K_1 \times K_2 = K\) element composite vector of explanatory variables consisting of the elements of \(X_1\) and \(A\), and \(\beta\) is the composite \((K \times 1)\) vector of parameters. This specifi-
Figure 3-1: The Latent Variable Model Specification with Performance Maintenance Interactions
cation is shown in Figure (3-2).

3.3 Some Special Cases of Linear Latent Variable Models:

Special cases of the linear model specification of the previous section can be obtained by making different restrictive assumptions on the MIMIC model. Many of these special cases are well known econometric models.

A: Reduced Form Models

As discussed in Chapter 2, one of the methods used by highway agencies in estimating deterioration models is to estimate a separate model for the propagation of each individual damage measurement. These models can be incorporated into the general linear latent variable framework as follows:

Substituting equation (3-5a) in (3-5b), we get:

\[ I = \lambda \beta X + \lambda \eta_1 + \epsilon = \pi X + \mu \]  

(3 - 6)

where \( \pi = \lambda \beta \) and \( \mu = \lambda \eta_1 + \epsilon \).

Equations (3-6) are a \((M \times 1)\) system of equations that link each measured damage type to the exogenous variables, and correspond to the independent damage models described above. These models do not rely upon the composite indices, since in equation (3-6), both the dependent and the independent variable are observed. Such a model is a "reduced form" of the model of equation (3-5).

One reduced form model is estimated for each damage type (or even for each severity level of each damage type). The practical problem with this method is that there is no systematic
Figure 3-2: Specification of a MIMIC Model
procedure for deciding which models to choose, since there is no obvious method for evaluating the different measurements making up the reduced form model system. In the latent variable system, the evaluation of the relevance of a measurement can be made on the basis of how much the measurement contributes to the unobserved condition. The measurement of this contribution is made on the basis of a goodness of fit type statistic called the "squared multiple correlation", which is analogous to the $R^2$ statistic obtained from a linear regression model. As explained in greater detail in Chapter 4, the total variance of an indicator can be broken up as the sum of two components: a component that is a measure of the variance of the latent variables that is "explained" by the indicator and a component $\theta$ which is simply the variance of the measurement error $\epsilon$ of the indicator. The contribution of an indicator in "explaining" a latent variable depends upon the relative magnitude of these two components. Those measurements which do not contribute very much to determination of the latent variable are characterized by a large measurement error variance $\theta$ values relative to the explained variance and can be discarded. An example of how this can be done in practice is provided in Chapter 6.

**B. Simultaneous Equation Models**

The traditional econometric simultaneous equation formulation can also be specified as a special case of the linear latent variable model. Consider equations (3-4). If $\lambda$ is an identity matrix, and the measurement error $\epsilon$ is zero, then equations (3-4c) reduce to the following:

$$I = S \tag{3-7}$$

Equations (3-4) can then be written as follows:

$$I = \beta_1 A + \beta_2 X_1 + \eta_1 \tag{3-8a}$$
\[ A = y + \Gamma X_2 + \eta_2 \]  \hspace{1cm} (3-8b)

In equation (3-8a) and (3-8b), all the variables are observed, and the measurement model does not exist any more. This is a traditional simultaneous equation specification.

C. Errors in Variable Models

If the explanatory variables in equation (3-5) are measured with error \( \delta \), then the following equations can be added to the measurement model:

\[ \hat{X} = X + \delta \]  \hspace{1cm} (3-9)

where \( \hat{X} \) is the measured value of \( X \). In this case, the measurement errors are due to errors in the measurement devices. This corresponds to an errors-in-variables formulation for the explanatory variables. This model can be combined with the simultaneous equation model of the previous sub-section.

D. Factor Analysis Model

If \( \beta, \eta_1 \) are zero, then the only equations that remain are the measurement model equations (3-5b). This corresponds to a model where the value of the latent variables are extracted from observed measurements of damage, with no associated deterioration relationship. Such a model is called a factor analysis model, because the correlations between the observed variables in the data are explained by a smaller number of latent "factors". The factor analysis model by itself is a widely used model to investigate the nature of observed correlations in the data.

3.4: Effects of Different Error Correlation Patterns:
In addition to constraints on the parameter matrices, assumptions about the nature of deterioration and about the manner in which the damage data is collected can also give rise to different specifications, since these assumptions correspond to different patterns of correlations between the error terms of the structural equations and the measurement equations. Ideally, since deterioration is temporal in nature, and since damage data is collected over time, the most logical model specification would be to have one set of structural and measurement equations for each time period. The possible correlation patterns between the error terms then correspond to the following scenarios:

a) Correlation between the structural equation error terms across time periods: The error term \( \eta_i \) in equation (3-4) accounts for unknown factors that affect deterioration in any time period. An example of such factors could be unknown effects of construction quality. If two time periods are spaced close together, they might both be subject to the same weather conditions. The error terms would then be correlated.

b) Correlation between error terms of different structural equations in the same time period \( t \) (i.e. correlation between \( \eta_{1j}(t) \), \( \eta_{1k}(t) \), \( j \neq k \) where \( j \) and \( k \) refer to two structural equations): This correlation can occur if two different processes of deterioration can be influenced by the same exogenous variable; for example, both load related and environment related deterioration can be affected by traffic. If there is a measurement error in the traffic variable, then this error affects the error terms of both the load related and the climate related latent performance variable.

c) Correlation between the measurement errors of different damage types in the same time period (i.e. correlation between \( \varepsilon_l(t) \) and \( \varepsilon_m(t) \) in eqn (3-5a) where \( l \) and \( m \) are subscripts of different measurements equations and \( t \) is the time period): This correlation occurs when the
same measuring device is used to measure more than one damage type. Suppose damage data is collected through visual inspections for a highway pavement. If the same inspector inspects a pavement surface for both longitudinal and alligator cracking, for example, then any inherent characteristics that the inspector might have (for example, a particular convention for measuring crack lengths) will appear in the error terms of the measurement equations for both damage types.

d) Correlation between measurement errors across time periods: If the same inspector inspects the facility in two time periods, then for the same reasons as in the previous case, the error terms of the measurement equation will be correlated across time periods.

From the above discussion, it is apparent that the latent variable model framework offers substantial flexibility in modeling many different deterioration mechanisms and measurement situations and is a general formulation covering many of the traditional econometric specifications.

3.5 Identification of Latent Variable Models

The linear latent variable model specification of equations (3-5) consist of two components; 1) a simultaneous structural equation system that consists of the deterioration and the maintenance extent equations and b) the measurement equation system. Two separate issues of identification are considered in this section. The first issue deals with the identification of the structural model, when empirical data is collected from in-service pavements. The second issue is the identification of the entire model system involving both the structural and the measurement models.
For a discussion of the first issue, consider the structural model specified in equations (3-4a) and (3-4b). In order to analyze the identification of this system, let us assume that the condition of the pavement is not latent, but is represented by some single known performance index, such as the PSI. Also assume that the vector of explanatory variables $X_1$ for the deterioration equation has only two elements: average daily traffic (ADT) and age, and that the explanatory variable vector $X_2$ for the maintenance equation has a single element: age. Also assume that the vector $A$ measures the extent of a single activity: crack filling. Equations (3-5a) and (3-5b) are then a system of two simultaneous equations that can be written as shown below:

$$\text{PSI} = \beta_{21} ADT + \beta_{22} Age + \beta_1 A + \eta_1 \quad (3 - 10a)$$

$$A = \gamma_1 \text{PSI} + \Gamma_1 Age + \eta_2 \quad (3 - 10b)$$

The reduced form of equations (3-10) can be obtained by substituting for $A$ and PSI respectively in the right hand sides of equations (3-10a) and (3-10b). The reduced form equations can be written as follows:

$$(1 - \beta_1 \gamma_1)\text{PSI} = (\beta_1 \Gamma_1 + \beta_{22}) Age + \beta_{21} ADT + (\eta_1 + \beta_1 \eta_2) \quad (3 - 11a)$$

$$(1 - \beta_1 \gamma_1)A = (\Gamma_1 + \gamma_1 \beta_{22}) Age + (\gamma_1 \beta_{21}) ADT + (\gamma_1 \eta_1 + \eta_2) \quad (3 - 11b)$$

Equations (3-11a) and (3-11b) can be rewritten as follows:

$$\text{PSI} = \Pi_{11} Age + \Pi_{12} ADT + \mu_1 \quad (3 - 12a)$$

$$A = \Pi_{21} Age + \Pi_{22} ADT + \mu_2 \quad (3 - 12b)$$

where $\Pi_{11} = (\beta_1 \Gamma_1 + \beta_{22})/(1 - \beta_1 \gamma_1)$ and the other composite terms are similarly defined.
Also,
\[ \eta_1 = \mu_1 - \beta_1 \mu_2 \]  \hspace{1cm} (3 - 12c)
\[ \eta_2 = \mu_2 - \gamma_1 \eta_1 \]  \hspace{1cm} (3 - 12d)

If equations (3-12a) and (3-12b) are estimated separately using ordinary least squares or maximum likelihood, four parameter estimates, for \( \Pi_{11}, \Pi_{12}, \Pi_{21}, \Pi_{22} \) are obtained. These four parameters are expressed in terms of the five parameters of equation (3-10), i.e. \( \beta_{21}, \beta_{22}, \beta_1, \gamma_1, \Gamma_1 \). Therefore, it is not possible to estimate all the parameters of equation (3-10).

The fact that the necessary condition is not satisfied does not mean that none of the parameters are estimable. The parameters that can be estimated can be derived by substituting equations (3-12) into (3-10). The following equations are then obtained:

\[ \Pi_{11} = \beta_{22} + \beta_1 \Pi_{21} \]  \hspace{1cm} (3 - 13a)
\[ \Pi_{12} = \beta_{21} + \beta_1 \Pi_{22} \]  \hspace{1cm} (3 - 13b)
\[ \Pi_{21} = \Gamma_1 + \gamma_1 \Pi_{11} \]  \hspace{1cm} (3 - 13c)
\[ \Pi_{22} = \gamma_1 \Pi_{12} \]  \hspace{1cm} (3 - 13d)

From equation (3-13d), \( \gamma_1 = \Pi_{22}/\Pi_{12} \) and \( \Gamma_1 = \Pi_{21} - \gamma_1 \Pi_{11} \). The two parameters for the maintenance equation can therefore be solved for. However, only two equations exist for the three parameters of the deterioration equation (3-11), and so these parameters are not identified. The maintenance equation is identified and the deterioration equation is not.

The identification problem is demonstrated pictorially in Figure (3-3). Since the reduced form equations (3-12) are obtained by solving for the endogenous variables in the structural equation
1 DETERIORATION RELATIONSHIP
   Condition = f(Maintenance)

2 MAINTENANCE RELATIONSHIP
   Maintenance = f(Condition)

Figure 3-3: Simultaneity of Deterioration and Maintenance
system (3-10), these equations simply describe the **point of intersection** of the deterioration equation (3-10a) and the maintenance equation (3-10b). Each observation on a pavement provides one such intersection point. The objective of the estimation procedure is to determine the slopes of the maintenance and the deterioration equations from these intersection points. In general, it is not possible to uniquely identify the slope of two lines simply from their intersection points, since there are an infinite number of lines that intersect at any given point. However, in the system of equation (3-10), there is enough information about the pattern of intersection points that allow the determination of one equation. In equation (3-10a), pavement condition depends upon traffic and age. If there is variability in the data, then different intersection points between the deterioration and maintenance equations are observed for different values of age. For each value of age therefore, a deterioration equations can be plotted as shown in Fig (3-3), where the lines labeled 1 are the deterioration equation and the line labeled 2 is the maintenance model. The maintenance equation can therefore be traced out from these different deterioration lines as shown in Fig(3-3). Since there is no explanatory variable on which a similar exercise can be carried out with the maintenance equation, the deterioration equation cannot be identified in this fashion. In order for both equations to be identified, each equation in the simultaneous equation system should contain at least one explanatory variable not found in the other equations. This is a simplified statement of the well known order condition for identification [Kmenta, 1980].

Traditional deterioration models specify only a **single** equation for the system of equation (3-10). If the deterioration equation is not identified, then estimation of a single equation could actually result in the estimation of the maintenance equation or some uninterpretable combination of the two equations, since not all parameters are estimable. The estimated "deterioration" model therefore is actually the wrong model with unpredictable parameter signs. As discussed in Chapter 1, the problem of incorrect parameter signs reported in the deterioration
literature could simply be the case of incorrect model estimation. An indication of this is provided in the case study described in Chapter 5. Estimation of a single deterioration equation produced a model where the condition improved with increasing traffic loads. With a proper simultaneous equation specification, however, the model exhibited the expected parameter signs.

In order to identify the deterioration model as well, one additional restriction is required in the model system that will generate one more equation. This restriction can be obtained by assuming that \( E(\eta_1, \eta_2) \) is zero. This assumption implies that the error terms of the structural and the maintenance equations are uncorrelated. The physical processes that cause deterioration and trigger maintenance activities are different, so there is no reason why the two error terms should be correlated. Making this assumption produces the following additional equation:

\[
E(\eta_1, \eta_2) = E(\mu_1 - \beta_1 \mu_2) (\mu_2 - \gamma_1 \mu_1) \text{ from equations (3-12c) and (3-12d).}
\]

Therefore, \( E(\eta_1, \eta_2) = 0 \) implies:

\[
(1 + \beta_1 \gamma_1) \mu_{12} - \gamma_1 \mu_{11} - \beta_1 \mu_{22} = 0 \quad (3-13e)
\]

where \( \mu_{12} = E(\mu_1 \mu_2) \), \( \mu_{11} = E(\mu_1^2) \) and \( \mu_{22} = E(\mu_2^2) \). These values can be obtained from the residuals after OLS estimation of equations (3-12) is performed.

Equations (3-13a), (3-13b), (3-13c), (3-13d) and (3-13e) together form a system of five equations from which the five unknown parameters can be estimated. The system is now exactly identified. The covariance restriction, equation (3-13e), however is different from the other equations in that it is non-linear. Therefore, theoretically, multiple roots to the equation
system are possible, giving rise to multiple solutions for the parameter estimates. In this case, however, equation (3-13e) involves only $\beta_i$ and $\gamma_i$, and once $\gamma_i$ is determined from equation (3-13d), $\beta_i$ is determined uniquely. The same should hold for the case where the parameters of the maintenance equation are over-identified. For each possible solution to $\gamma_i$, a unique solution will exist for $\beta_i$. The possibility of multiple solutions does not therefore appear to be a problem in this specification.

The above discussion concentrated on the identification of the structural equation system, assuming that there was no latent variable in the model. When latent variables are present, the joint identification of both the structural and the measurement model system should be considered. The problem with the estimation of latent variable models is that, except in a few specific cases, no general sufficient condition exists to test whether the model is identified or not. For clarity, the identification issues will be demonstrated for the MIMIC model without the simultaneous structural equation specification. The concepts are directly applicable to the augmented model as well.

Consider the MIMIC specification of equation (3-5) repeated below:

$$S = \beta X + \eta_1$$  \hfill (3-5a)

$$I = \lambda S + \varepsilon$$  \hfill (3-5b)

In equation (3-13), $S$ is a $(P \times 1)$ latent variable vector, $X$ is a $(K \times 1)$ vector and $I$ is a $(M \times 1)$ vector. $\beta$ is then a $(K \times 1)$ vector of parameters, and $\lambda$ is a $(M \times 1)$ vector. Typically, for estimation, the following assumptions are made about the model:
1) \( E(\varepsilon) = 0, E(\eta_1) = 0 \)

2) \( E(\varepsilon \eta_1') = 0 \)

3) \( \varepsilon \) is uncorrelated with \( S \)

4) \( \eta_1 \) is uncorrelated with \( X \).

Assumption (2) assumes that the measurement error terms of the indicator variables measuring performance are uncorrelated with the specification error terms of the structural equation. This is a reasonable assumption since the deterioration process and the measurement processes are independent of each other. Note that no assumptions have been made about the covariance matrices of \( \varepsilon \) or \( \eta_1 \). If the same inspector measures all the damage components on a pavement, then any measurement biases that the inspector introduces into the measurement will be manifested in the errors of all measurements. In this case, \( E(\varepsilon \varepsilon') \) will not be diagonal. Similarly, if pavement performance can be characterized by more than one latent variable that share common deterioration mechanisms, the structural equation error terms will be correlated with each other. Assumption (3) states that the measurement errors and the latent variables are uncorrelated. Since the latent variables are theoretical constructions to explain observed phenomena, they can be constructed without loss of generality to be independent of the measurement errors. Assumption (4) is similar to that made in a linear regression model; since the error terms account for items not covered by the explanatory variables, the explanatory variables and the error terms are uncorrelated.

As described in detail in Chapter 4, the parameters of the model of equation (3-5) are estimated by fitting the sample covariance matrix calculated from the observed variables (\( I \) and \( X \)) in the data to a calculated covariance matrix implied by the model which can be expressed as a function of the parameters. For the specification of eqn (3-5), the calculated covariance matrix \( \Sigma \) between the observed variables can be written as follows:
\[
\Sigma(.) = \begin{pmatrix}
\Sigma_{II} & \Sigma_{IX} \\
\Sigma_{IX} & \Sigma_{XX}
\end{pmatrix}
\]

(3 - 14a)

and the observed (sample) variance covariance matrix \( V(.) \):

\[
V(.) = \begin{pmatrix}
V_{II} & V_{IX} \\
V_{IX} & V_{XX}
\end{pmatrix}
\]

(3 - 14b)

This partitioned variance covariance matrix of equation (3-14) are both \((K+M) \times (K+M)\) matrices. With assumptions 1,2 and 3 above, the submatrices of equation (3-14a) can be expressed as follows in terms of the parameters:

\[
\Sigma_{II} = E((\lambda S + \epsilon) [\lambda S + \epsilon]')
\]

\[
= E[\lambda S S']\lambda' + E[\lambda S \epsilon'] + E[\epsilon S']\lambda' + E[\epsilon \epsilon']
\]

(3 - 15)

where \( E[.]. \) stands for expected value. By assumption 3, the two middle terms of equation (3-15) are zero, and equation (3-15) becomes:

\[
\Sigma_{II} = \lambda E(SS')\lambda' + \theta
\]

(3 - 16)

where \( \theta \) is the variance covariance matrix of the error term \( \epsilon \).

The variance covariance matrix of \( S \) can be obtained from eqn (3-14a) as follows:
\[ E(\delta S') = E[(\beta X + \eta_1)(\beta X + \eta_1)] = \]
\[ \beta \Phi \beta' + \Psi \]  
(3-17)

where \( \Phi \) is the variance covariance matrix of \( X \) and \( \Psi \) is the variance of the error \( \eta_1 \). By assumption (4), the two cross product terms in equation (3-17) are zero.

From equations (3-16) and (3-17),

\[ \Sigma_{\eta} = \lambda (\beta \Phi \beta' + \Psi) \lambda' + \theta \]  
(3-18)

Equation (3-18) expresses the top left hand corner of the partitioned variance covariance matrix \( \Sigma \) of equation (3-14a) in terms of the parameter vectors. The two other components of \( \Sigma \) can be similarly derived to give:

\[ \Sigma_{xx} = \Phi \]  
(3-19)
\[ \Sigma_{\eta x} = \lambda \beta \Phi \]  
(3-20)

In equations (3-18), (3-19) and (3-20), the elements of \( \Sigma_{\eta}, \Sigma_{xx} \) and \( \Sigma_{\eta x} \) are functions of the elements of \( \lambda, \beta, \Phi, \theta \) and \( \Psi \).

The number of non duplicatable elements in the observed variance covariance matrix of equation (3-14b) is \((K+M)(K+M+1)/2\) since there are \( M \) indicators for the latent variable and \( K \) explanatory variables. The number of parameters to be estimated from the model of equation (3-5) can be tallied as follows:
a) $K$ parameters of $\beta$

b) $M$ parameters of $\lambda$

c) 1 parameter $\psi$, the variance of the structural equation error term.

d) $M(M+1)/2$ parameters of $\Theta$, the covariance matrix of measurement errors.

e) $K(K+1)/2$ parameters of $\Phi$, the covariance matrix of the explanatory variables.

If the total number of parameters to be estimated (i.e., $K + M + 1 + M(M+1)/2 + K(K+1)/2$) is greater than the number of observed variances and covariances $(K+M)(K+M+1)/2$, then some parameters of the five matrices above have to be set to zero. Usually, the nature of the problem automatically defines these constraints; for example, in the specification of equation (4-1), $\Phi$, is a fixed matrix, and the problem in this case reduces to estimating $(K + M + 1 + M(M+1)/2)$ parameters from $(K+M)(K+M+1)/2 - K(K+1)/2 = M(M+1)/2 + MK$ observed variances and covariances. The necessary condition in this case is $MK > M+K+1$. If there is no reason to assume that the correlations between the different performance indicators, then $\Theta$ is diagonal, and only $M$ parameters need to be estimated for $\Theta$ rather than $M(M+1)/2$. In this fashion, parameters can be constrained to satisfy the necessary condition.

Unfortunately, the fact that a specification satisfies the necessary condition does not mean that the model is identified. For example, consider the situation described above where the necessary identification condition is $MK > M + K + 1$. For $M>2$, $K>2$, this condition is always satisfied without further restrictions. This does not necessarily mean that the model is identified. One means of checking identification described by Joreskog [1973, 1979] and others is to examine in detail each equation that links an observed covariance to a calculated covariance, and check to see if each parameter can be estimated uniquely. Often, the equations are complex, and this is not easy to do in practice. An example of how such an analysis can be carried out is now described.
Consider data collected on highway pavement sections on four different damage types that are used as indicators for a single latent performance variable \( S \). These indicators could be, for example, for each section:

1) Extent of high severity alligator cracking in sq. ft. \( (I_1) \)
2) Extent of medium severity alligator cracking in sq. ft. \( (I_2) \)
3) Extent of high severity rutting (in.) \( (I_3) \)
4) Extent of medium severity rutting (in.) \( (I_4) \)

In the case of alligator cracking, severity is measured in terms of crack widths, and for rutting, severity depends upon the area of pavement section that is rutted (i.e. upon the rut widths). It may be unlikely that medium severity and high severity rutting are found on the same pavement section since rutting is usually concentrated on the wheel paths, but we can assume that they coexist for purposes of illustration. These four measurements could be indicators for a "load related" performance variable. Suppose also that data is available on two explanatory variables: age of the section and ADT. The following model is specified from this data:

A) Structural equation:

\[
S = \beta_{21} ADT + \beta_{22} \log e + \eta_1
\]  
\[
(3-21a)
\]

B) Measurement equations:

\[
I_1 = \lambda_1 S + \epsilon_1
\]  
\[
(3-21b)
\]

\[
I_2 = \lambda_2 S + \epsilon_2
\]  
\[
(3-21c)
\]
\[ I_3 = \lambda_3 S + \epsilon_3 \quad (3-21d) \]
\[ I_4 = \lambda_4 S + \epsilon_4 \quad (3-21e) \]

The variables that are observed in the above model are the four damage measurements and the two explanatory variables. The observed variance covariance matrix \( V \) is therefore a \((6 \times 6)\) matrix, with \((6 \times 7)/2 = 21\) unduplicated terms. The elements of the matrix \( V \) are denoted as \( V_{ij}, i=1,2,3,4,5,6 \) and \( j=1,2,3,4,5,6 \). The first four rows of \( V \) correspond to the indicators, and the last two to the explanatory variables.

The \((6 \times 6)\) calculated variance covariance matrix \( \Sigma \), can be partitioned as follows from equations \((3-18), (3-19)\) and \((3-20)\):

\[
\Sigma(.) = \begin{pmatrix}
\lambda \psi \lambda' + \theta & \lambda \beta \Phi \\
\lambda \beta \Phi &  \Phi
\end{pmatrix} \quad (3-22)
\]

\( \theta \) is the variance covariance matrix of the error terms \( \epsilon \) and \( \psi \) is the variance of the latent variable. The bottom right hand corner of equation \((3-22)\) is the \((2 \times 2)\) variance covariance matrix of the explanatory variables \( X (= \Phi) \) with three independent elements. This matrix is fixed, since the explanatory variables are fixed and do not depend on any parameters. These three elements therefore do not contribute to the estimation of the parameters. There are therefore \((21-3) = 18\) remaining elements of \( \Sigma \) that are functions of the parameters, which can be compared with the eighteen corresponding terms of the observed covariance matrix. The number of parameters to be estimated are as follows:

a) 4 elements of \( \lambda \)
b) 2 elements of \( \beta \)
c) variance of $\eta = \psi$

d) $(4 \times 5)/2 = 10$ elements of $\theta_e$, the variance covariance matrix of $\varepsilon$

This gives a total of 17 parameters to be estimated from 18 equations, so the necessary conditions of identification are satisfied. However, despite the necessary conditions being satisfied, it is not possible to uniquely estimate all the parameters without further restrictions. This becomes apparent when the eighteen equations characterizing the system are written out. Suppose the covariance between $\varepsilon_i$ and $\varepsilon_j$ is written as $\theta_{ij}$. Equating corresponding unduplicated elements of the observed and calculated variance covariance matrix for $\Sigma_{\eta}$, the following ten equations are obtained:

\begin{align*}
\nu_{11} &= \lambda_1^2 \psi + \theta_{11} & (3-23a) \\
\nu_{12} &= \lambda_1 \lambda_2 \psi + \theta_{12} & (3-23b) \\
\nu_{13} &= \lambda_1 \lambda_3 \psi + \theta_{13} & (3-23c) \\
\nu_{14} &= \lambda_1 \lambda_4 \psi + \theta_{14} & (3-23d) \\
\nu_{22} &= \lambda_2^2 \psi + \theta_{22} & (3-23e) \\
\nu_{23} &= \lambda_2 \lambda_3 \psi + \theta_{23} & (3-23f) \\
\nu_{24} &= \lambda_2 \lambda_4 \psi + \theta_{24} & (3-23g) \\
\nu_{33} &= \lambda_3^2 \psi + \theta_{33} & (3-23h) \\
\nu_{34} &= \lambda_3 \lambda_4 \psi + \theta_{34} & (3-23i) \\
\nu_{44} &= \lambda_4^2 \psi + \theta_{44} & (3-23j)
\end{align*}

The ten equations (3-23a to 3-23j) express the 10 observed variances and covariances in terms of the elements of the parameter matrices $\lambda$, $\psi$ and $\theta$. Equations can be similarly written for
the eight parameters of $\Sigma_{xx}$. These equations correspond to observed values of $v_{ij}$, where i=5,6 and j=5,6. The following eight equations are obtained:

\[
\begin{align*}
\nu_{15} &= \lambda_1 \beta_{21} \phi_{11} + \lambda_1 \beta_{22} \phi_{12} \\
\nu_{25} &= \lambda_2 \beta_{21} \phi_{11} + \lambda_2 \beta_{22} \phi_{12} \\
\nu_{35} &= \lambda_3 \beta_{21} \phi_{11} + \lambda_3 \beta_{22} \phi_{12} \\
\nu_{45} &= \lambda_4 \beta_{21} \phi_{11} + \lambda_4 \beta_{22} \phi_{12} \\
\nu_{16} &= \lambda_1 \beta_{21} \phi_{12} + \lambda_1 \beta_{22} \phi_{22} \\
\nu_{26} &= \lambda_2 \beta_{21} \phi_{12} + \lambda_2 \beta_{22} \phi_{22} \\
\nu_{36} &= \lambda_3 \beta_{21} \phi_{12} + \lambda_3 \beta_{22} \phi_{22} \\
\nu_{46} &= \lambda_4 \beta_{21} \phi_{12} + \lambda_4 \beta_{22} \phi_{22}
\end{align*}
\]

(3 - 24a) - (3 - 24h)

$\phi_{11}, \phi_{12}, \phi_{22}$ are the three fixed elements of the covariance matrix of the explanatory variables. On examining the set of equations (3-23) and (3-24), it is seen that equations (3-24) are a set of eight equations expressed in terms of only six parameters; there are therefore multiple ways of solving for some of the parameters of equation (3-24). On the other hand, equations (3-23) are a set of ten equations expressed in terms of fifteen parameters, which means that the measurement model system is not identified. Even if $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ were calculated from equation (3-24) and substituted back in equation (3-23), there are still only 10 equations, but eleven parameters to be estimated. Therefore, though the total set of equations (3-23) and (3-24) satisfy the necessary condition, the fact that there is a redundancy in equations (3-24) and a deficit in equations (3-23) makes some of the measurement model parameters unidentified.

So far, no constraints have been placed on the parameters in the model specification. Some
constraints are now necessary to identify the model. Clearly, the nature of the constraints depends upon the specification and the circumstances of the problem being modeled. For example, the unconstrained specification has assumed that all the measurement errors ε are correlated with each other. If the four measurements of the indicators are made independently, then it can be argued that there is no basis for the correlation between the error terms of the measurement equation, and the six covariance (i.e. off diagonal) terms of the matrix Θ can be set to zero. There are now only 9 parameters (4 values of λ, ψ, 4 values of θ) to be estimated from 10 measurement equations, and this makes the measurement model over-identified as well. The system of equations can now be written as follows:

\[\begin{align*}
\nu_{11} &= \lambda_1^2 \psi + \theta_{11} \\
\nu_{12} &= \lambda_1 \lambda_2 \psi \\
\nu_{13} &= \lambda_1 \lambda_3 \psi \\
\nu_{14} &= \lambda_1 \lambda_4 \psi \\
\nu_{22} &= \lambda_2^2 \psi + \theta_{22} \\
\nu_{23} &= \lambda_2 \lambda_3 \psi \\
\nu_{24} &= \lambda_2 \lambda_4 \psi \\
\nu_{33} &= \lambda_3^2 \psi + \theta_{33} \\
\nu_{34} &= \lambda_3 \lambda_4 \psi \\
\nu_{44} &= \lambda_4^2 \psi + \theta_{44}
\end{align*}\]

\( (3 - 25a) \)

\( (3 - 25b) \)

\( (3 - 25c) \)

\( (3 - 25d) \)

\( (3 - 25e) \)

\( (3 - 25f) \)

\( (3 - 25g) \)

\( (3 - 25h) \)

\( (3 - 25i) \)

\( (3 - 25j) \)

Despite the fact that the measurement model now satisfies the necessary condition for identification, the model system is still not identified. One additional restriction is needed in order to fix the units of measurement for the latent variable. Recall that of the two damage types
serving as indicators for the load related performance variable, alligator cracking extent is measured in terms of an area and rutting extent in terms of a rut depth. This means that there is no natural units in which the latent variable S can be defined, and the scale of S is therefore arbitrary. To illustrate this, consider equation (3-25b). For a given value of \( \psi \) an infinite number of values of \( \lambda_1 \) and \( \lambda_2 \) satisfy the equation. This is because for any non zero \( "m" \), \( m\lambda_1 \) and \( \lambda_2/m \) give the same value of \( \psi_{12} \). One of the methods of fixing the scale of the model is done by constraining the value of one element in each column of the parameter matrix \( \lambda \). All the other parameter values are then estimated relative to these values. It is usual practice to set the fixed values of \( \lambda \) to be equal to one. The units in which the latent variable is measured is then the same as the units of the corresponding indicator. Assume \( \lambda_1 = 1 \). The latent variable then has the units of sq. ft.\{ the units of alligator cracking\}.

With the scale of the latent variable also defined, the parameters of the measurement model may now be obtained as follows from equation (3-25):

\[
\lambda_3 = \psi_{23}/\psi_{12} \text{ from eqns f and b.} \\
\lambda_5 = \psi_{23}/\psi_{13} \text{ from eqns f and c.} \\
\lambda_4 = \psi_{13}/\psi_{12} \text{ from eqns i and c.} \\
\psi = \psi_{12}/\psi_{13}/\psi_{23} \text{ from eqns b,f and c.} \\
\theta_{11} = \psi_{11} - (\psi_{13}/\psi_{12}/\psi_{23}) \text{ from eqn a.} \\
\theta_{22} = \psi_{22} - (\psi_{23}/\psi_{12}/\psi_{13}) \text{ from eqn e.} \\
\theta_{33} = \psi_{33} - (\psi_{23}/\psi_{12}/\psi_{13}) \text{ from eqn h.} \\
\theta_{44} = \psi_{44} - (\psi_{24}/\psi_{12}/(\psi_{13}/\psi_{23})) \text{ from eqn j.} \\
\]

The over identification in the above model is indicated by the fact that equations (3-25d) and (3-25g) are never used. These equations can be used to give other estimates for the parameters;
for example, \( \lambda_3 \) can be expressed using \( d \) and \( i \) as \( \nu_{3d}/\nu_{14} \). In addition, multiple estimates of \( \lambda \) and therefore of \( \beta \) can be obtained from the over-identified structural equations.

The problem with the method described above for checking model identification is that the number of equations increases rapidly as the number of parameters to be estimated increases. The method therefore is not suitable for analyzing complex specifications involving a large number of parameters.

The above analysis was performed for the specification of equation (3-5) in which the simultaneity of deterioration and maintenance was not explicitly formulated. Identification for the maintenance equations can be checked in a similar fashion. As was seen in the above example, the structural model was over-identified since the eight equations (3-24) are expressed in terms of six parameters. This is typical for a MIMIC model. Therefore, in general, the complete set of equations (deterioration and maintenance) will be over-identified with respect to the measurement model as long as these equations are identified with respect to each other.

Once the identification of the model has been checked, the next step is to estimate the parameters of the latent variable model. Estimation methods are described in Chapter 4.
CHAPTER 4
ESTIMATION OF LATENT VARIABLE MODELS

In this chapter issues relating to the estimation and goodness-of-fit tests for the latent variable model formulated in Chapter 3 are discussed. The specification that will be used for the purposes of illustration in this chapter is the MIMIC model of equations (3-5) with a single latent variable S, a vector of explanatory variables X, and a vector I of indicators for the latent variable. Equations (3-5) are repeated below:

\[ S = \beta X + \eta_1 \]  
\[ I = \lambda S + \epsilon \]

where S is a latent variable, I is a vector of indicators, X is a vector of explanatory variables including maintenance, \( \eta_1 \) and \( \epsilon \) are structural and measurement equation error terms respectively. Denote \( E(\eta_1 \eta_1') \) as \( \psi \) and \( E(\epsilon \epsilon') \) as \( \theta \). If only a single latent variable exists, then clearly there is only a single structural equation.

4.1 Estimation of a Latent Variable Model

4.1.1 Observed Covariance Matrix as Function of Parameters

The estimation of a latent variable model requires obtaining parameter estimates for the vectors \( \beta, \lambda \) and the parameters of the covariance matrices \( \psi \) and \( \theta \) of the error terms \( \eta \) and \( \epsilon \). The estimation is carried out by calculating the population variance covariance matrix for the
observed variable vectors $X$ and $I$ of equation (4-1). The form of the specification of equation (4-1) implies a particular form for the population variance covariance matrix, since the elements of this matrix are functions of the parameters. The parameters are estimated by minimizing a function of the discrepancy between $\Sigma(.)$ and $V$, where $\Sigma(.)$ is the calculated variance-covariance matrix implied by the model and $V$ is the usual unbiased sample variance covariance matrix obtained from a sample of $N$ observations on the observed variables.

Consider the model specified in equation (4-1), where $S$ is a ($P \times 1$) latent variable vector, $X$ is a ($K \times 1$) vector and $I$ is a ($M \times 1$) vector. $\beta$ is then a ($M \times 1$) vector of parameters, and $\lambda$ is a ($K \times 1$) vector. In this case, each "observation" from 1 to $N$ contains information on $K$ explanatory variables and $M$ indicators. The set of equations (4-1) therefore applies to each observation.

If the specification of equation (4-1) is subject to the assumptions discussed in section (3-5), then from equations (3-18), (3-19) and (3-20), the calculated covariance matrix $\Sigma(.)$ can be expressed as follows:

$$\Sigma(.) = \begin{pmatrix} \lambda \psi \lambda' + \theta & \lambda \beta \Phi \\ \lambda \beta \Phi & \Phi \end{pmatrix}$$

(4 - 2)

From equation (4-2), the elements of $\Sigma(.)$ are elements of the parameter matrices $\lambda$, $\beta$, $\Phi$, $\theta$ and $\psi$.

In any particular application, some of the elements of the five parameter matrices mentioned
above will be fixed (usually to zero), some will be unknown but constrained to equal other elements, and some will be unknown and unconstrained. For example, \( \Phi \), is a matrix of variances and covariances of the observed explanatory variables \( X \). A consistent estimator for \( \Phi \) is the sample covariance matrix for the explanatory variables calculated from the data. Therefore, the elements of \( \Phi \) are fixed and are set equal to the corresponding sample values. Similarly, \( \theta \) is the covariance matrix of the measurement model error terms \( \varepsilon \). If the error terms of the causal model are not correlated across equations, then \( \theta \) is a diagonal matrix and its off-diagonal elements are set to zero. An example where parameter values are set equal to each other arises in the case of a covariance structure model formulated over \( T \) time periods, with a separate structural and measurement model for each period. Since the parameter values of the measurement equations do not change across time periods, the elements of the matrices \( \lambda \) are constrained to be equal for each time period.

Summarizing, the elements of the parameter matrices are of three kinds:

1. fixed parameters which have been assigned given values
2. constrained parameters whose values are unknown but are equal to one or more other parameters
3. free parameters which are unknown and not constrained to be equal to any other parameter.

As discussed in Section 3.5, the total number of independent covariances in the observed covariance matrix \( V \) is \((K+M)(K+M+1)/2\). The total number of free parameters must not be greater than this in order to satisfy the necessary conditions for identification. Other restrictions may be necessary if the model is still not identified.
4.1.2 Estimation of Parameters

The method for the estimation of parameters for a latent variable model is to choose a vector of parameters in such a manner as to minimize a "discrepancy function" $D(V; \hat{\Sigma})$. This function is a scalar valued function which gives an indication of the discrepancy between the sample variance covariance matrix $V$ and the calculated matrix $\hat{\Sigma}$ for any particular value of the parameters. The function has the following properties [Everitt, 1984]:

1) $D(V; \hat{\Sigma}) \geq 0$
2) $D(V; \hat{\Sigma}) = 0$ if and only if $\hat{\Sigma} = V$
3) $D(V; \hat{\Sigma})$ is continuous in $V$ and $\hat{\Sigma}$

The discrepancy between the empirical variance covariance matrix and the calculated matrix can be written as follows:

$$V = \hat{\Sigma}(.) + \Delta$$  \hspace{1cm} (4 - 3)

where $\Delta$ is a $(K+M \times K+M)$ matrix of discrepancies between $V$ and $\hat{\Sigma}$. The simplest estimator uses a least squares discrepancy function, which can be written as follows:

1) **Ordinary least squares:**

The discrepancy function is:
where \( \text{vech}(\Delta) \) indicates that the \((K+M)(K+M+1)/2\) independent elements of the matrix \( \Delta \) have been stacked to produce a column vector.

As can be seen, the ordinary least squares criterion satisfies the properties of a discrepancy function listed above. The criterion has a major problem: it is not independent of the scale of the manifest variable. In other words, different estimates of the parameter matrix are produced if the measured variables undergo any linear transformation involving multiplication by a diagonal non-identity matrix. Also, all deviations between the observed and the calculated covariance matrices are weighted equally. In practice, it is more desirable to weight the deviations so that the variables that are measured more accurately get a higher weight. This leads to the following generalized least squares criterion:

2) **Asymptotic Distribution Free (ADF) Estimator:**

In this case, the discrepancy function is:

\[
D(V, \hat{\Sigma}(\cdot)) = \text{vech}(\Delta')W^{-1}\text{vech}(\Delta)
\]  

(4 - 5)

where \( W \) is a positive definite matrix of weights of dimension equal to the number of elements in \( \Delta \). Browne [1978] shows that any estimate \( \hat{\delta}_{ADF} \) of the parameter matrix obtained by minimizing a discrepancy function with properties 1), 2) and 3) is consistent. The estimator of equation
(4-5) is distribution free because the objective function of equation (4-5) is simply a generalized likelihood function criterion: no specific distributional assumption is made. Hsiao[1982] shows that the minimum variance GLS estimator can be found when $W$ is an estimate of the covariance matrix of $\Delta$. Since equation (4-5) is expressed in terms of variances and covariances, a typical element of $W$ will consist of fourth order moments and products of second order moments.

3) Normal Theory Minimum Distance (NMDE) Estimator:

If the observed variables have a multivariate normal distribution, or any other distribution where kurtosis is not present, then the off-diagonal product terms of $W$ are zero, and the discrepancy function given by eqn (4-5) is arithmetically equivalent to the following more computationally efficient expression [Browne, 1974]:

$$F(V, \hat{\Sigma}(\cdot) \mid W) = 0.5 \text{trace}[\Delta'W^{-1}\Delta]$$  \hspace{1cm} (4 - 6)

where $W$ is any consistent estimator of the variance covariance matrix of $\Delta$. One obvious choice for $W$ in equation (4-6) is the sample variance covariance matrix $V$. Equation (4-6) then becomes:

$$F(V, \hat{\Sigma}(\cdot) \mid W = V) = 0.5 \text{trace}[\Delta'V^{-1}\Delta]$$  \hspace{1cm} (4 - 7)

It is easily seen that the function given by eqn (4-7) satisfies properties (1), (2) and (3). In addition, it is also scale free; if there is no restrictions on the parameter matrix, then the scale
of measurement does not affect the parameter values, since the least square function is scaled by the sample variance covariance matrix. Browne[1974] also shows that under the assumption of normality, the parameter estimates obtained by minimizing equation (4-7) are efficient.

4) **Maximum likelihood criterion:**

A fourth estimator can be obtained by assuming that the observed variables constitute a random sample drawn from a multivariate normal distribution. The sample variance covariance matrix \( V \) then has a **Wishart distribution**. The Wishart distribution is a multivariate extension of the univariate chi-squared distribution, and is defined as the sum of independent products of multivariate normal random vectors [Johnson and Wichern, 1982].

Under the assumption of multivariate normality of the observed variables, the maximum likelihood estimates can be obtained by maximizing the following log likelihood function for \( N \) observations each containing \( K \) explanatory variables and \( M \) indicators from a \( (K+M) \)-variate normal distribution [Lawley and Maxwell, 1971] with a function of the observations omitted:

\[
\ln L(\hat{\Sigma}(\cdot)) = -N/2[\ln | \hat{\Sigma}(\cdot) | + \text{trace}(V \hat{\Sigma}(\cdot)^{-1})]
\]

where \( | \Sigma | \) stands for the determinant of the calculated covariance matrix.

The maximum likelihood estimate for \( \hat{\Sigma}(\cdot) \) is the sample variance covariance matrix \( V \). When \( \hat{\Sigma}(\cdot) = V \), the value of the likelihood function is:
\[ \ln L(V) = -N/2[\ln |V| + \text{trace}(I_{K+M})] \]  \hspace{1cm} (4 - 9)

where \(I_{K+M}\) is an identity matrix with dimension equal to that of the observed variables.

Subtracting eqn (4-9) from (4-8), the following function is obtained:

\[ L(V, \hat{\Sigma}(\cdot)) = \ln L(V) - \ln L(\hat{\Sigma}(\cdot)) = \]
\[ = -N/2[\ln |\hat{\Sigma}(\cdot)| + \text{trace}(V\hat{\Sigma}(\cdot)^{-1}) - \ln |V| -(K+M)] \] \hspace{1cm} (4 - 10)

The function given by equation (4-10) has the following properties:

1) It is nonpositive, since eqn (4-8) is maximized at the value given by eqn (4-9).
2) It takes the value zero when \(\hat{\Sigma} = V\)
3) It is continuous in \(V\) and \(\hat{\Sigma}\)

Therefore, the negative of the function given by equation (4-10) satisfies the properties of a discrepancy function. The maximum likelihood estimates for the parameter matrix are then obtained by \textbf{minimizing} the following function:

\[ L(V, \hat{\Sigma}(\cdot)) = [\ln |\hat{\Sigma}(\cdot)| + \text{trace}(V\hat{\Sigma}(\cdot)^{-1}) - \ln |V| -(K+M)] \] \hspace{1cm} (4 - 11)

Let \(\alpha_{ML}^*\) be the vector of parameter values at which the function of equation (4-11) is minimized. Browne[1974] shows that if the value of the calculated variance covariance matrix \(\hat{\Sigma}(\alpha_{ML}^*)\) is
substituted for $V$ in the NMDE objective function of equation (4-7), i.e. $V = \Sigma(\alpha_{ML}^*)$, and the expression is minimized, the NMDE estimates will have the same value as the maximum likelihood estimates, i.e. eqn (4-7) and eqn (4-11) have their minima at the same point. The maximum likelihood estimator may then be regarded as a member of the asymptotically equivalent class of generalized least squares estimators with minimum asymptotic variance under the assumption of normality [Browne, 1982].

The NMDE and the MLE estimators presented above are the most commonly used for the estimation of parameters of a covariance structure where the observations are normally distributed or where no kurtosis is present. If kurtosis is present, (i.e., the multi-variate normality assumption fails) then use of a discrepancy function of the general form discussed in this section gives consistent parameter estimates but incorrect standard errors. Issues of deviations from normality will be discussed in a subsequent section.

4.2 Extraction of Latent Variable Scores

As described in Chapter 3, the objective of interest from a latent variable model is to extract the posterior distribution of the latent variable conditional on the distribution of the observed variables. Mostly, however, it is difficult to derive the actual distribution of the latent variable, but it is possible to extract the value of the latent variable from the observations. The value of the latent variable is extracted from the explanatory variables and the measured indicators using a method called the "regression method" [Johnson and Wichern, 1982]. This method arises from a generalization of a regression model that states that if there are two vectors of
random variables $S$ and $I$ that have finite second moments and a variance covariance matrix with elements $\Sigma_{II}, \Sigma_{IS}, \Sigma_{SS}$, then it is always possible to write the following relationship between $S$ and $I$:

$$S = \alpha_0 + \alpha_1 I + \nu$$

(4 - 12)

where $\alpha_1$ is a vector of parameters given by $\Sigma[II]^{-1}\Sigma[IS]$ and $E(\nu)=0$. The expression for $\alpha_1$, which is the covariance between $S$ and $I$ multiplied by the inverse of the variance of $I$ is the same as the least squares estimator for the parameters of a linear regression model. This is why the method for the extraction of the latent variable gets its name.

In order to derive expressions for $\alpha_0$ and $\alpha_1$, the specification for the linear latent variable model is repeated from Chapter 3:

$$S = \beta_1 A + \beta_2 X_1 + \eta_1$$

(4 - 13a)

$$A = \gamma S + \Gamma X_2 + \eta_2$$

(4 - 13b)

$$I = \lambda S + \varepsilon$$

(4 - 13c)

where $I$ is a vector of indicators, $S$ is the latent variable, $X_1$ is a vector of deterioration equation explanatory variables, $A$ is a vector of extents of maintenance activities and $X_2$ is the vector of explanatory variables for the maintenance models.

The first step in the extraction of the latent variable is to express the simultaneous equation structure of equation (4-13) in its reduced form. This can be done by substituting each of the
maintenance equations into the deterioration equation, so that $S$ is expressed as the function of $X_1$ and $X_2$. The following equation is obtained:

$$S = \beta_1 \gamma S + \beta_1 \Gamma X_2 + \beta_2 X_1 + \eta_1 + \beta_1 \eta_2$$  \hspace{1cm} (4 - 14)

This can be rewritten as follows:

$$S = B \tilde{X} + \tilde{\eta}$$  \hspace{1cm} (4 - 15)

where $\tilde{X}$ consists of $X_1$ and $X_2$, $B$ is a composite vector comprising $\{(1 - \beta_1 \gamma)^{-1} \beta_2, (1 - \beta_1 \gamma)^{-1} \beta_1 \Gamma\}$ and $\tilde{\eta}$ is a composite error term consisting of $\{\eta_1, \beta_1 \eta_2\}$.

From equation (4-12), $\alpha_0$ is given by:

$$E(S | I, \tilde{X}) = \alpha_t E(I)$$  \hspace{1cm} (4 - 16)

From equations (4-13c),

$$E(S | I, \tilde{X}) = B \tilde{X}$$  \hspace{1cm} (4 - 17)

From equations (4-13c),
\[ E(I) = \mathcal{E}(S) \]
\[ \lambda B \tilde{X} \quad (4 - 18) \]

As derived in (3-18), the variance covariance matrix of the indicators \( \Sigma_{\eta} \) is given by:

\[ \lambda \tilde{\psi} \lambda' + \theta \quad (4 - 19) \]

where \( \tilde{\psi} = E(\tilde{\eta} \tilde{\eta}') \) and \( \theta = E(\varepsilon \varepsilon') \). Since the \( X \)'s are fixed, the variance of \( S \) is equal to \( \tilde{\psi} \) and the covariance between \( I \) and \( S \) is \( \lambda \tilde{\psi} \). \( \alpha_i \) is then given by:

\[ \lambda \tilde{\psi}(\lambda \tilde{\psi} \lambda' + \theta)^{-1} \quad (4 - 20) \]

and \( \alpha_0 \) by:

\[ B \tilde{X} - \alpha_i \lambda B \tilde{X} \quad (4 - 19) \]

From (4-20) and (4-21), the extracted value of \( S \), \( \hat{S} \) is given by:

\[ \hat{S} = B \tilde{X} + \lambda \tilde{\psi}(\lambda \tilde{\psi} \lambda' + \theta)^{-1}(I - \lambda B \tilde{X}) \quad (4 - 22) \]

Equation (4-22) can be used to calculate the value of the latent variable for each observation from corresponding observations of the latent variable and the indicators.

4.3 Goodness-of-fit Tests
These tests determine how well the specified model fits the observed data. If the parameters are estimated by minimizing any of the fitting functions described in Section 4.1, a measure of the overall fit of the model is given by the \((N - 1)\) times the minimum value of the discrepancy function, where \(N\) is the sample size. This function is approximately chi-squared distributed with degrees of freedom equal to the total number of observed variances and covariances minus the number of estimated parameters.

The chi-squared statistic compares the goodness of fit of the specified model with respect to a model on which no constraints are imposed, i.e. in which none of the parameters are pre-specified to be zero. The total number of parameters estimated is equal to the total number that can be estimated, i.e. the total number of observed variances and covariances. The point at which the objective function for the unconstrained model is minimized is clearly when the calculated covariance matrix duplicates the observed sample matrix, and at this point the function value and the \(\chi^2\) are both zero. The constrained specification is one where some of the parameters are constrained to be zero. Consider a specification where the total number of parameters that can be estimated from the unconstrained model is \(T\), and the parameters are labeled \(a_1, a_2, \ldots, a_T\). In the constrained model, suppose without loss of generality that the first \(R\) parameters are set to be zero. Then the null hypothesis tested by the chi squared statistic is:

\[
H_0: a_1 = a_2 = a_3 = \ldots = a_R = 0
\]  

(4-23)

If the chi squared statistic from a constrained model is small enough (i.e. close enough to zero) then the hypothesis of equation (4-23) cannot be rejected, and the constrained model is
valid.

There are two major problems with the use of the chi-squared statistic as a measure of the goodness of fit for latent variable models. The first problem is its dependence on sample size. The statistic is both a function of sample size and the distance between the calculated and sample variance-covariance matrices. As a result, as the sample size increases, the probability of rejecting the constrained model increases, and if the sample size is large enough, almost all constrained models would be rejected on the basis of the test. A more reasonable manner in which to use chi-squared statistic is to compare differences between two constrained models, one of which is nested within the other. If the difference in the value of the statistic between the two models is large compared to the change in the degrees of freedom, then the additional parameters that were set to zero in the more restricted models contain information and cannot be discarded.

The second problem with the chi squared goodness of fit statistic is that is extremely sensitive to the effects of model misspecification. If, for example, the data does not come from a multivariate normal distribution, as is often the case in practice, then the statistic may no longer be asymptotically chi-squared distributed. In practice, therefore, researchers have found it very difficult to specify models that are not rejected by the chi-squared statistic. As a result, greater reliance has been placed on other, more ad-hoc statistics for evaluating the fit of the model [Joreskog and Sorbom, 1981]. A useful measure of fit of the individual equations of the model is a measure similar to the $R^2$ statistic in linear regression. This measure is called the squared multiple correlation (SMC) and is defined as follows:
For each structural equation $i$,

$$SMC_i = 1 - \frac{Var(\eta_i)}{Var(S_i)}$$  \hspace{1cm} (4-24a)$$

and for each measurement equation $j$,

$$SMC_j = 1 - \frac{Var(\epsilon_j)}{Var(I_j)}$$  \hspace{1cm} (4-24b)$$

where, as defined before, $S$ is the latent variable, $I$ is the indicator, $\eta$ is the structural equation error, and $\epsilon$ is the measurement equation error.

In the case study described in Chapter 5, the different model specifications are non-nested. In addition, many of the explanatory variables do not come from a multivariate normal distribution. The squared multiple correlation is therefore the primary statistic used to compare between different models in the case study.

4.4 Specification Issues

The maximum likelihood estimates of Section 4.1 and the chi-squared statistic described in Section 4.3 assume that the all the observed data is generated from a multivariate normal distribution. In practice, however, more often than not, this is not the case. For example, some
of the explanatory variables used in the deterioration equation are dummies that take discrete values, and are therefore not normally distributed. The multivariate joint distribution of the observed variables is therefore non-normal as well.

As discussed in Section 4.1, minimization of the Maximum Likelihood objective function produces maximum likelihood estimates when the data comes from a multivariate normal distribution. These estimates are asymptotically equivalent to the estimates produced using the generalized least squares objective function of equation (4-7). In this case, the parameter estimates are efficient as well. Browne [1982] demonstrates that the above properties are true not only for a normal distribution, but for any other distribution that is kurtosis-free. The kurtosis of a multivariate distribution is calculated from the Mardia coefficient of excess kurtosis [Mardia, 1974]. If \( W \) is a p-variate random variable, with mean given by the \((p \times 1)\) vector \( \bar{W} \) and variance covariance matrix \( \Sigma \), the Mardia coefficient of multivariate kurtosis is calculated from the following formula:

\[
E\{(W - \bar{W})\Sigma^{-1}(W - \bar{W})\}^2 - p(p + 2) \tag{4 - 25}
\]

If \( W \) has a multivariate normal distribution or any kurtosis free distribution (such as the truncated normal, or the multivariate \(t\) distribution), then the coefficient of multivariate kurtosis is zero, and the NMDE and ML estimators produce asymptotically equivalent efficient parameter estimates.

If the multivariate distribution that generates the data is kurtose, then the use of the Maximum likelihood objective function to obtain parameter estimates no longer produces Maximum Likelihood Estimates for the parameters. The use of the ML fitting function when the data is
not normal produces what are called "pseudo-maximum likelihood" estimates. If the parameter estimates are consistent, then it is shown in Gourieroux et. al. [1982] that the pseudo maximum likelihood estimator can be thought of as a generalization of the Maximum Likelihood estimator, where a different kind of information criterion is being maximized. As mentioned in Section 4.1, any estimates obtained by minimizing a discrepancy function of the ML or GLS type are consistent, so consistency of the parameter estimates is not a problem. However, since the estimators are not ML estimators any more, the variance covariance matrix of the parameters calculated in the traditional way by inverting the information matrix will be inconsistent. Bartholomew [1987] shows that when the distribution of the data is not multivariate normal, then the standard errors obtained from the information matrix underestimate the true standard errors. In order to obtain estimates for the variance covariance matrix that are "robust" to non-normality of the errors, a procedure suggested by White[1980] can be used. This procedure is a general method for obtaining consistent parameter estimates for the variance covariance matrix when a model is specified so that consistent but inefficient parameter estimates are produced. Using this procedure, the variance-covariance matrix of the parameters is the asymptotic variance covariance matrix of a class of estimators called "M (or Maximum Likelihood Type) Estimators" [Amemiya, 1985]. For a regression model written in the usual fashion (i.e., \( y = x\beta + u \)), the M-estimator minimizes some function of \((y - x\beta)/s\), where s is an estimate for the variance of u. This is a generalization of the Maximum-Likelihood estimator. White shows that under some conditions of boundedness of the error variance and the variance covariance matrix of the regressors, the following statistic:

\[
W = A^{-1}BA^{-1}
\]  

(4 – 26)
is a consistent estimate for the variance covariance matrix of the parameters $\beta$, where:

$$A = (1/N) \sum_{i=1}^{N} \frac{\partial^2 l_i}{\partial \beta \partial \beta'}$$  \hspace{1cm} (4-27a)

$$B = (1/N) \sum_{i=1}^{N} \left( \frac{\partial l_i}{\partial \beta} \right) \left( \frac{\partial l_i}{\partial \beta} \right)'$$  \hspace{1cm} (4-27b)

where $l_i$ is the likelihood function value (or pseudo-likelihood function value) for observation $i$. For each observation, the matrix $A$ is the information matrix, and the matrix $B$ calculates the expected value of the squares of the first derivatives. If the data does come from a multivariate normal distribution, then the information matrix is equal to the negative of the expected value of the square of the first order derivatives (i.e. $A=-B$) [White, 1982], and the variance covariance matrix of the parameters is simply $-A^{-1}$ which is the usual inverse of the negative of the information matrix. The expression for the variance-covariance matrix given by equation (4-26)

The amount of computation required in equation (4-24) is substantially more than that required for calculating the standard errors from the information matrix since equation (4-24) requires the computation of the first order derivative for each observation individually.

The assumption inherent in using the ML criterion for data from kurtose distributions is that the model is not misspecified to the extent that the parameter estimates are inconsistent. To test the consistency of the Maximum Likelihood parameters, a Hausman specification can be performed using the MLE and the ADF estimator of equation (4-5). Since the ADF estimator is simply a "minimize distance" criterion with the distance measure satisfying the assumptions
for a fitting function specified by Browne [1974], the parameter estimates obtained by this method are always consistent. If $\beta_{ML}$ is the vector of parameter estimates using the maximum likelihood estimator, $\beta_{ADF}$ the estimators using the GLS estimator, and $V_{ML}$ and $V_{ADF}$ their corresponding variance covariance matrices, then the following statistic:

$$ (\beta_{ML} - \beta_{ADF})' (V_{ML} - V_{ADF})^{-1} (\beta_{ML} - \beta_{ADF}) $$

(4 - 28)

is chi-squared distributed with degrees of freedom equal to the number of estimated parameters [Hausman, 1978]. If the parameters using Maximum Likelihood are consistent, then the chi-squared statistic of equation (4-26) will not be significantly different from zero. If the hypothesis that there is no significant difference between the estimators is rejected, then the ADF estimator should be used because it is distribution free and is more intuitively appealing.
CHAPTER 5
DESCRIPTION OF RESULTS

In this chapter, a case study is described that implements the latent variable model on data collected in the field. Several models are presented starting from simple models that duplicate the specifications of past studies and leading up to more realistic models of pavement performance.

5.1 Description of Data

The data set used for the case study was collected by ARE [Butler, Carmichael and Flanagan, 1985] from the state of Nevada. The standard unit of observation is characterized as a "rural, flexible, full unit mile pavement section". The information contained in the data set for each year and each pavement section can be summarized as follows:

a) Damage Measurements: Square feet of alligator and linear cracking, widths of alligator and linear cracking, square feet of patching, average rut depth and slope variance.

b) Environment and traffic variables: ADT, daily truck traffic, age of section since last rehabilitation, minimum temperature, annual precipitation, number of freeze/thaw cycles and number of wet days.

c) Maintenance activities: Extent of maintenance performed on 10 activities, the number of labor hours and the total material cost.

The data was collected for 4871 pavement sections over a period of 5 years from 1980 through 1984. Each year several teams of inspectors drove the entire system evaluating each mile for damage. The procedure was to examine the first two-thirds of the mile and then select a typical
1000 sq ft. section in the last third of the mile to evaluate in detail. The condition of these 1000 sq. ft. is taken as representative of the condition of the section as a whole. The resulting damage measurements reflect different 1000 sq. ft. sections each year.

The Nevada data set was chosen because it is one of the few available data sets that has substantial maintenance information which is critical for the development of models predicting deteriorative. In the analysis performed by ARE with the data, four "preventive maintenance" activities were isolated from among the ten activities available in the data set. These were the activities that were performed regularly on the pavement, and in amounts exceeding more than a few square yards.

The obvious deterioration model that can be specified with time series data links the performance of the facility at some time t to a sequence of all events that occurred in the past. For example, the condition of a highway pavement at time t can be linked to the traffic in past years, to the maintenance performed, and to the condition before and after maintenance in the past. Such a specification would entail the definition of a latent variable for each performance characteristic for each year, and indicators of the performance characteristics for each year. However, in order to estimate such a model it is necessary to know the exact sequence of past events such as: at what point in the year the maintenance was performed, what the condition was before and after maintenance and so on. At a minimum, for each year in which a routine maintenance activity is performed, it would be necessary to know what was done since the inspection or the maintenance activity. It would make sense to assume that the annual inspection always precedes the maintenance activity for the year. However, the elapsed time between the most recent maintenance activity and the subsequent inspection can only be grossly estimated. Conversations with the Nevada Department of Transportation revealed that it was not possible
to extract this necessary information from the data set. This made it difficult to specify a time series model from the Nevada data, and so a sectional model was estimated with the data used in the same fashion as was used by ARE.

As a result, the five years of data had to be pooled together for each pavement section. The data was averaged over the five years to produce average values of damage, condition, maintenance and the explanatory variables over a five year period. The five year averages represent the best method for representing the long term effects of maintenance; on the average, a better maintained pavement section will be in a better condition. Rather than predict the condition of the pavement every year, the models presented in this section predict the average condition over a five year period. This information will be useful to highway agencies for medium to long range maintenance planning.

The ARE study used the PSI as the index of pavement performance. The PSI is calculated from the following equation: [Butler, Carmichael and Flanagan, 1985]

\[
PSI = 5.03 - 1.91 \log_{10}(1 + SV) - 1.38RD^2 - 0.03(C + P)^{0.5}
\]

(5-1)

where SV is the slope variance in units of in², RD is the rut depth in inches and C and P are extents of cracking and patching in square feet respectively.

The results described later in this section will use the PSI model as the basis of comparison.

A large number of the 4871 pavement sections had errors, miscodings and missing data; cleaning of the data and removal of outliers brought the number of observations down to 3837. The variables used in the analysis for each section are summarized in Table (5-1). Each variable is a five year average; for variables that are non-linear functions of other variables (such as the
a) average PSI multiplied by (-1); measure of pavement condition (-PSI)
b) average sq. yds. of sand seal coat per year (SANDSEAL)
c) average lbs. of filler used for crack filling per year (CR. FILL.)
d) average sq. yds. of chip seal coat per year (CHIPSEAL)
e) average sq. yds. of flush seal coat per year (FLUSHSEAL)
f) average daily traffic (ADT)
g) average daily truck traffic as percentage of ADT (TRUCKS)
h) average age of section (AGE)
i) square of average age (AGE2)
j) district 1 dummy (DIST1)
k) district 2 dummy (DIST2)
l) road surface type dummy (SURF. TYPE)
  0 = flexible overlay on rigid pavement
  1 = flexible overlay on flexible pavement
m) average annual precipitation (in.) (PRCPTN)
n) average annual number of freeze-thaw cycles (FRZE/THAW)
o) average annual minimum temperature (MIN TEMP)
p) average annual number of days with precipitation (WETDAYS)
q) average extent of linear or alligator cracking on section (sq. ft.) (C)
r) average extent of patching on section (sq. ft.) (P)
s) average rut depth on section (inches) (RD)
t) average slope variance on section (in²) (SV)
u) average width of linear cracking on section (in.) (Al. Width)
v) average width of alligator cracking on section (in.) (Lin. Width)

TABLE 5-1: Description of Variables used in Analysis
(Reference names in parentheses)
PSI), the value of the variable was calculated for each year and then averaged over the five years. For example, the average daily truck traffic as a percentage of ADT was calculated by computing the percentage truck traffic for each of the five years and then taking their average.

The ranges and mean values of the data are summarized in Table 5-2(a), and the coefficient of univariate kurtosis is reported in Table 5-2(b) for some selected variables. The reason for including data on the kurtosis is that, following the discussion of Section 4.4, the traditional method for calculating standard errors for the parameter estimates is incorrect if the data comes from a substantially kurtose distribution. In this case, the method suggested by White [1982] in Section 4.4 needs to be used. Note that the PSI variable used in all the following models is actually the negative of the PSI and increasing values of this index reflect increased deterioration of the pavement. The data was collected from three districts, labeled districts 1, 2 and 3. Some summary traffic and climate statistics are presented in Table 5-2(c) for each district (for the 3837 observations). Conversations with the Nevada Highway Administration revealed that district 1 is in the south of Nevada, with a very dry climate. The major arterial highway through Nevada goes through this district, so there is higher car and truck traffic in district 1 than in the other districts. Districts 2 and 3 are in the mountains with more precipitation than district 1, but with less traffic. Analysis of the data, however, indicates that there are no substantial differences between the traffic in district 1 and district 2, and the district 2 traffic even appears to be higher.

The fact that some of the high traffic roads are in areas that are environmentally less susceptible to deterioration creates an ambiguity in the role that traffic plays in deterioration. This ambiguity is compounded by the fact that overall traffic volumes, except in a few heavily travelled sections, are very low. 50% of the pavement sections have volumes of less than 200 vehicles per day, and 90% of the sections have less than 2000 vehicles per day. In addition, as can be
<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSI</td>
<td>-2.55</td>
<td>-1.25</td>
<td>4.14</td>
</tr>
<tr>
<td>ADT</td>
<td>690.00</td>
<td>5.00</td>
<td>8990.00</td>
</tr>
<tr>
<td>NO. OF TRUCKS</td>
<td>126.00</td>
<td>0.00</td>
<td>1680.00</td>
</tr>
<tr>
<td>AGE</td>
<td>14.40</td>
<td>3.00</td>
<td>18.00</td>
</tr>
<tr>
<td>AGE2</td>
<td>231.89</td>
<td>9.00</td>
<td>324.00</td>
</tr>
<tr>
<td>MIN. TEMP. (degrees F)</td>
<td>36.30</td>
<td>20.00</td>
<td>56.00</td>
</tr>
<tr>
<td>FREEZE-THAW CYCLES</td>
<td>147.00</td>
<td>0.00</td>
<td>230.00</td>
</tr>
<tr>
<td>PRECIPITATION (in.)</td>
<td>8.50</td>
<td>2.00</td>
<td>35.00</td>
</tr>
<tr>
<td>WETDAYS</td>
<td>44.20</td>
<td>11.00</td>
<td>81.00</td>
</tr>
<tr>
<td>AREA OF SAND SEALCOAT (yds²)</td>
<td>1600.00</td>
<td>0.00</td>
<td>7250.00</td>
</tr>
<tr>
<td>AMOUNT OF CRACK FILLING (lbs.)</td>
<td>807.00</td>
<td>0.44</td>
<td>5300.00</td>
</tr>
<tr>
<td>AREA OF CHIP SEALCOAT (yds²)</td>
<td>570.00</td>
<td>0.00</td>
<td>8017.00</td>
</tr>
<tr>
<td>(CRACKING + PATCHING)⁰⁵ (ft.)</td>
<td>11.68</td>
<td>0.00</td>
<td>35.10</td>
</tr>
<tr>
<td>(RUT DEPTH)² (in²)</td>
<td>0.02</td>
<td>0.00</td>
<td>0.50</td>
</tr>
<tr>
<td>log₁₀ (1 + SLOPE VARIANCE)</td>
<td>0.92</td>
<td>0.00</td>
<td>1.59</td>
</tr>
</tbody>
</table>

Table 5-2(a): Summary Statistics of Variables
<table>
<thead>
<tr>
<th>Variable</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>-PSI</td>
<td>2.49</td>
</tr>
<tr>
<td>AREA OF SAND SEAL COAT</td>
<td>7.12</td>
</tr>
<tr>
<td>AMOUNT OF CRACK FILLING</td>
<td>18.55</td>
</tr>
<tr>
<td>AREA OF CHIP SEAL COAT</td>
<td>7.76</td>
</tr>
<tr>
<td>(CRACKING + PATCHING)^{0.5}</td>
<td>2.47</td>
</tr>
<tr>
<td>(RUT DEPTH)^2</td>
<td>31.20</td>
</tr>
<tr>
<td>(\log_{10}(1 + \text{SLOPE VARIANCE}))</td>
<td>2.57</td>
</tr>
</tbody>
</table>

Table 5-2(b): Kurtosis of Selected Variables

<table>
<thead>
<tr>
<th>District</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>average daily traffic</td>
<td>765</td>
<td>1050</td>
<td>350</td>
</tr>
<tr>
<td>average daily truck traffic</td>
<td>125</td>
<td>170</td>
<td>90</td>
</tr>
<tr>
<td>average precipitation(in./yr)</td>
<td>6.5</td>
<td>8.0</td>
<td>10.5</td>
</tr>
<tr>
<td>average annual min. temperature</td>
<td>42</td>
<td>35.2</td>
<td>31.9</td>
</tr>
</tbody>
</table>

Table 5-2(c): Mean Values of Key Variables by District
seen from Table 5-2(a), even the areas that have rainfall receive relatively small amounts of precipitation; the average precipitation is only 8.5 inches, and the average number of wet days (i.e. days with any precipitation at all) is only 44, which corresponds to less than four days of precipitation a month. A more suitable data set for deterioration modeling would be one with substantial traffic loads and varying weather patterns; unfortunately a data set of this nature was not available.

5.2 A Single Equation PSI Model

Table 5-3 shows the simplest model that can be estimated on the data set, referred to as MODEL0. This is a single equation ordinary least squares (OLS) regression model with the negative of PSI as the dependent variable. This is the type of model which was previously estimated for this data set by ARE. Note that all the variables are in deviations form (i.e. for each observation, each variable has been subtracted from it's sample mean). The reason for this is to facilitate comparison with the latent variable models. Since the latent variable model parameters are estimated by the comparison of variance covariance matrices, the models are specified with the observed variables expressed as deviations from their mean values, and it is not possible to estimate intercepts. However, it is not required to estimate the intercepts. The reason for this is that the deterioration index (both the PSI and the latent variable) are measured on an arbitrary scale. There is no "natural" physical interpretation for the intercept, and so even if it were estimated, it can be absorbed by rescaling the performance index.

Four preventive maintenance activities performed on the pavements are: 1) sand seal coat 2) flush seal coat 3) chip seal coat and 4) crack filling. Since there are very few observations on which flush seal coating has been performed, this activity was not used in the analysis. The
Dependent Variable: (-PSI)

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Estimate</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Maintenance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sand Seal Coat Area</td>
<td>-0.151</td>
<td>-6.74</td>
</tr>
<tr>
<td>Crack Filling Amount</td>
<td>0.316</td>
<td>3.72</td>
</tr>
<tr>
<td>Chip Seal Coat Area</td>
<td>-0.179</td>
<td>-5.94</td>
</tr>
<tr>
<td><strong>Traffic and Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Av. Daily Traffic</td>
<td>-1.550</td>
<td>-4.12</td>
</tr>
<tr>
<td>Percentage Trucks</td>
<td>2.110</td>
<td>5.11</td>
</tr>
<tr>
<td>Age</td>
<td>1.447</td>
<td>8.98</td>
</tr>
<tr>
<td>Age2</td>
<td>-0.142</td>
<td>-6.64</td>
</tr>
<tr>
<td><strong>Environmental</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Av. Annual Precipitation</td>
<td>-0.255</td>
<td>-2.59</td>
</tr>
<tr>
<td>Av. Annual Min. Temp.</td>
<td>-0.064</td>
<td>-0.82</td>
</tr>
<tr>
<td>Av. Annual Freeze/Thaw Cycles</td>
<td>-0.223</td>
<td>-2.39</td>
</tr>
<tr>
<td>Av. No. of Wet days</td>
<td>0.675</td>
<td>5.52</td>
</tr>
<tr>
<td><strong>Dummies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District 1 Dummy</td>
<td>-0.173</td>
<td>-1.06</td>
</tr>
<tr>
<td>District 2 Dummy</td>
<td>0.283</td>
<td>2.02</td>
</tr>
<tr>
<td>Road Surface Dummy</td>
<td>2.334</td>
<td>23.16</td>
</tr>
</tbody>
</table>

No. of observations: 3837
R²: 0.38

Table 5-3: Deterioration Model Parameters: MODEL0
three seal coating activities are water proofing activities. The objectives for performing these activities, as stated by a pavement management manual from California quoted by Humplick[1984] are as follows:

a) Waterproof pavement
b) Prevent Crack Spalling
c) Stop Raveling
d) Restore Binder Flexibility

The service life of these activities is 1-3 years if properly used. Proper usage of these activities is on pavements where the sealing has dried out, and some fine aggregate raveling has taken place. None of the seal coating activities are recommended for use if the ravel is coarse, or there are heat cracks present; in addition, chip sealing which uses rubberized asphalt chips, should not be performed if substantial cracking is present without first filling the cracks. Crack filling is also a waterproofing activity used on a clean crack more than 1/4" wide; it should not be used on dirty cracks or on cracks less than 1/4" wide.

The parameters of MODEL0 were estimated using the negative of PSI divided by 1.91 as the dependent variable. The units of PSI are then expressed in the units of log(slope variance), which are the same units in which the latent performance variables of the specifications described later. As can be seen from Table 5-3, many significant parameter estimates have the wrong signs. For example, the model shows that the condition of the pavement deteriorates as the extent of crack filling increases. Similarly, traffic has the wrong sign with the pavement condition improving as the ADT increases. Some environmental variables have the wrong sign as well. The two most significant variables are the road surface type dummy and age. The road surface type dummy has a value of one if the road surface is concrete mix, and a value
of 0 if the surface is bituminous mix. The effect of age is specified as a polynomial function that involves both age and the square of age. Finally, the district dummy parameters are not significant: with district 3 as the reference, pavements in district 1 have somewhat higher PSI's, while those in district 2 have somewhat lower PSI's. On looking at the table of summary statistics for each district, district 2 has the highest traffic and more precipitation than district 1. District 3 has more precipitation than the other two sections, but substantially less traffic. This might be the reason for the difference between the districts.

MODEL0 is similar to the models estimated by ARE on the same data set. ARE found the same problems that MODEL0 has, i.e., unexpected signs on important explanatory variables and poor overall fit. ARE concluded that the data set contained too many measurement errors to successfully estimate a model. While this might be true, it is also likely that the reason for the failure of MODEL0 is that it fails to recognize the simultaneity between the deterioration process and the agency maintenance behavior models. As discussed in Section 2.6, a proper specification of pavement deterioration models from empirical data involves the simultaneous estimation of two models: the first representing deterioration as a function of traffic, age, maintenance and climate, and the second representing the amount of maintenance performed as a function of the factors affecting maintenance. What these factors are may be difficult to specify without knowing the exact nature of the maintenance practice of the agency; however, traffic is clearly a factor that influences maintenance behavior, with high volume roads receiving more maintenance. If there is no variable in the maintenance equation that does not appear in the deterioration equation, then the deterioration equation is not identified. However, the agency behavior model is likely to be overidentified, since there are many exogenous variables (e.g. the environmental variables) that appear in the deterioration equation but not in the agency behavior equation. Therefore, an attempt to estimate a "deterioration" model by a single equation as in MODEL0 might actually be estimating the maintenance behavior of the agency.
rather than the deterioration relationship, which might explain the incorrect sign of the traffic parameter. The correct specification therefore is one where the effect of simultaneity between maintenance and deterioration is explicitly included in the formulation. The results obtained from such a specification are presented in the following section.

5.3 A Simultaneous Equation PSI Model

The simultaneous equation specification is an extension of MODEL0. The deterioration model is identical to the specification of MODEL0 (with -PSI as the dependent variable, and the independent variables listed in Table 5.3). In addition, three maintenance equations are specified, one each for chip sealing, sand sealing and crack filling. The specification can be written as follows:

\[-\text{PSI} = \beta_1 A_1 + \beta_2 A_2 + \beta_3 A_3 + \beta_4 X_1 + \eta_1 \quad (5-2a)\]

\[A_1 = \gamma_1 (-\text{PSI}) + \Gamma_1 X_2 + \eta_2 \quad (5-2b)\]

\[A_2 = \gamma_2 (-\text{PSI}) + \Gamma_2 X_2 + \eta_3 \quad (5-2c)\]

\[A_3 = \gamma_3 (-\text{PSI}) + \Gamma_3 X_1 + \eta_4 \quad (5-2d)\]

where \(X_1\) is the 11 element vector of explanatory variables from Table (5-3) that includes all the listed variables except for the three maintenance variables, \(X_2\) is a vector of explanatory variables for the maintenance models consisting of the following:

1) Average Daily Traffic
2) Percentage of Trucks
3) Road Surface Type.
\( A_1, A_2 \) and \( A_3 \) respectively are the amounts of sand sealing, crack filling and chip sealing performed on the pavement and \( \Gamma_1, \Gamma_2, \Gamma_3 \) are each \((3 \times 1)\) vectors of parameters.

Equation \((5-2a)\) specifies the deterioration model, while the other three equations specify the maintenance model for each of the three routine maintenance activities considered. As discussed in the previous section, traffic, expressed both in terms of ADT and the percentage of trucks is an important determinant of maintenance. The second important criterion that determines the amount of maintenance performed is clearly the condition of the pavement; the better the condition, the higher the maintenance performed. Finally, the amount of maintenance performed is likely to differ by the road surface type. The specification of the explanatory variables for the maintenance model was arrived at after conversations with the Nevada Highway Administration, who confirmed that these were the major variables upon which their maintenance decisions were based. Note that MODEL1 is still not a latent variable model; all the variables are observed.

As discussed extensively in Chapter 3, identification of the structural equation system requires that each equation of the system contain some explanatory variables that are not included in the other equations. In the model system the maintenance equation \((5-2b), (5-2c)\) and \((5-2d)\) do not include a number of exogenous variables that are included in \((5-2a)\). The information on these variables can be used to trace out the maintenance equation curves as shown in Fig (3-3). The maintenance equations are therefore identified with respect to the deterioration equation.

However, without further restrictions, the deterioration equation in the above system is not identified, since there are no exogenous variables in equations \((5-2b), (5-2c)\) and \((5-2d)\) that do not appear in equation \((5-2a)\). One additional restriction needs to be placed between the
structural equation and each maintenance equation. As discussed in Chapter 3, the restrictions that $E(\eta_1, \eta_2), E(\eta_1, \eta_3), E(\eta_1, \eta_4)$ are all zero satisfy the necessary conditions for the identification of (5-2a).

The maximum likelihood estimates of the parameters of MODEL1 are tabulated in Tables 5-4(a) and 5-4(b). As mentioned earlier, the results are normalised by using the PSI divided by 1.91 (the coefficient of the log(slope variance) term in the PSI equation (5-1)) as the dependent variable to make the results comparable with subsequent models. Table 5-4(a) shows the deterioration equation parameters, and Table 5-4(b) the parameters of the maintenance equations. As is apparent from these tables, all the parameters have the expected sign. From Table 5-4(a), the condition of the pavement improves as the maintenance increases, and gets worse as the traffic and the percentage of trucks increase, and as the pavement gets older. The significant environmental variables also have the expected signs. As in MODEL0, the age function and the road surface dummy are the most significant variables. Traffic and most of the environmental effects except for the number of wetdays are all not significant. This is a reasonable result since in the data set, traffic levels are low, and there is not much precipitation. In Table 5-4(b) as well, all parameters have the expected signs. Maintenance increases as the pavement condition gets worse, reflected by the positive sign of the coefficient of condition in the maintenance equation. Similarly, the maintenance activity increases as traffic and the percent of trucks increase.

The overall fit of each equation of the model system is measured by the "squared multiple correlation" which is defined for each equation as one minus the ratio of the variance of the error term $\eta$ to the variance of the dependent variable. The squared multiple correlation is the same as the $R^2$ for this model. For MODEL1, the squared multiple correlation for the structural equation is 0.27, from Table 5-4(a). This value is less than that obtained from MODEL0, which
**Dependent Variable: (-PSI)**

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Estimate</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Maintenance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sand Seal Coat Area</td>
<td>-0.434</td>
<td>-3.01</td>
</tr>
<tr>
<td>Crack Filling Amount</td>
<td>-1.676</td>
<td>-6.46</td>
</tr>
<tr>
<td>Chip Seal Coat Area</td>
<td>-0.042</td>
<td>-0.16</td>
</tr>
<tr>
<td><strong>Traffic and Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Av. Daily Traffic</td>
<td>0.067</td>
<td>0.76</td>
</tr>
<tr>
<td>Percentage Trucks</td>
<td>2.649</td>
<td>6.65</td>
</tr>
<tr>
<td>Age</td>
<td>1.743</td>
<td>10.83</td>
</tr>
<tr>
<td>Age2</td>
<td>-0.172</td>
<td>-5.46</td>
</tr>
<tr>
<td><strong>Environmental</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Av. Annual Precipitation</td>
<td>-0.205</td>
<td>-1.43</td>
</tr>
<tr>
<td>Av. Annual Min. Temp.</td>
<td>-0.141</td>
<td>-1.72</td>
</tr>
<tr>
<td>Av. Annual Freeze/Thaw Cycles</td>
<td>-0.308</td>
<td>-3.30</td>
</tr>
<tr>
<td>Av. No. of Wet days</td>
<td>0.453</td>
<td>3.30</td>
</tr>
<tr>
<td><strong>Dummies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District 1 Dummy</td>
<td>-0.618</td>
<td>-2.54</td>
</tr>
<tr>
<td>District 2 Dummy</td>
<td>0.209</td>
<td>0.60</td>
</tr>
<tr>
<td>Road Surface Dummy</td>
<td>1.700</td>
<td>13.58</td>
</tr>
</tbody>
</table>

No. of observations: 3837  
R²: 0.28

Table 5-4(a): Deterioration Model Parameters: MODEL1

**Dependent Variable: Maintenance Activity**

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Sand Seal Coat Estimate/t-statistic</th>
<th>Chip Seal Coat Estimate/t-statistic</th>
<th>Crack Filling Estimate/t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>-PSI</td>
<td>0.154/2.40</td>
<td>-0.049/-0.75</td>
<td>0.072/7.83/</td>
</tr>
<tr>
<td>Av. Daily Traffic</td>
<td>0.147/3.67</td>
<td>0.047/1.25</td>
<td>0.113/8.30</td>
</tr>
<tr>
<td>Percentage Trucks</td>
<td>2.122/6.13</td>
<td>0.853/2.84</td>
<td>-0.235/-3.08</td>
</tr>
<tr>
<td>Road Surface Dummy</td>
<td>-1.471/-7.64</td>
<td>0.532/2.74</td>
<td>-0.311/-9.89</td>
</tr>
</tbody>
</table>

R²: 0.05  0.05  0.03

Table 5-4(b): Maintenance Model Parameters: MODEL1
is only to be expected since MODEL1 is a simultaneous equation formulation of the single regression equation of MODEL0. A single equation OLS would yield the highest possible value of $R^2$ for that equation. The simultaneous equation estimator gets rid of the bias, but in the process worsens the fit. In some situations one may accept some level of bias to achieve a better fit. However, in this case MODEL0 is unacceptable because key parameters were estimated to have counter-intuitive signs. The next step, therefore, is to specify a model that retains the simultaneous equation formulation, but improves the fit. This can be done by introducing latent variables into the formulation.

5.4 A PSI type Latent Variable Model

MODEL2 is an extension to the formulation of MODEL1 that includes a latent variable. Instead of having (-PSI) as the dependent variable in the specification of equation (5-2a), it is now assumed that the dependent variable is some latent variable S representing the condition of the pavement. The structural equation system remains the same as before except that (-PSI) is replaced in each equation by the latent variable S. The explanatory variables are assumed to be measured without error, so no measurement equations need to be specified for these variables. However, a measurement model is needed for the indicators of the latent performance variable. The idea behind the choice of the latent variable for this model is that it should measure the same underlying performance characteristic that the PSI does; therefore, it is necessary that the latent variable be constructed from the same damage components that make up the PSI. The three measurements used as indicators for the latent variable are therefore the square root of cracking and patching, the square of the rut depth and the logarithm of the slope variance. The measurement equations are specified as follows:

$$(C + P)^{0.5} = \lambda_i S + \varepsilon_i$$

(5-3a)
\[ RD^2 = \lambda_2 S + \varepsilon_2 \quad (5-3b) \]

\[ \log_{10}(1 + SV) = \lambda_3 S + \varepsilon_3 \quad (5-3c) \]

The maximum likelihood estimates for the parameters of the structural model and the measurement model of MODEL2 are shown in Table 5-5. Table 5-2(b) indicates that some of the endogenous variables have substantial kurtosis. As mentioned in Section 4.4, if there is substantial kurtosis in the data, the maximum likelihood estimator gives consistent but inefficient parameter estimates, since the data is not generated from a multivariate normal distribution. One method of testing the specification of the model is to compare the parameter estimates obtained from the Maximum Likelihood estimator to the parameter estimates obtained from a distribution free estimator such as the asymptotic distribution free (ADF) estimator described in Section (4.4) using the Hausman test statistic. The Hausman test, as described in Section 4.4 involves estimating the parameters of the model using two estimators, one of which is known to produce consistent parameter estimates. If the parameter estimates produced by the estimator to be tested converges to the estimates produced by the consistent estimator, then the test estimator is also consistent. The convergence is tested by a \( \chi^2 \) distributed statistic which is constructed from the difference between the parameter estimates obtained by the two estimators (see equation 4-31). Unfortunately, the software required for the estimation of parameters using the ADF estimator was not available; the Hausman test was therefore performed using the Normal Theory Minimum Distance Estimator (NMDE) as the alternate estimator to the MLE. The problem with the use of the NMDE estimator that both the MLE and NMDE are based upon assumptions of multivariate normality of the error terms. If the Hausman test indicates a similarity in the parameter estimates obtained from the two estimators, then this is an indication that the model is well specified. However, if the test indicates a dissimilarity (as evidenced by the value of the \( \chi^2 \) statistic being significantly different from zero) between the estimates produced by the two estimators, then the results are difficult to
interpret. This is because a divergence in the parameter estimates does not indicate which estimator (if any) is mis-specified. Fortunately, in this case, a Hausman test indicated that the difference between the structural parameters from the ML and the NMDE estimators was not significant. The standard errors for the parameters are calculated by the method described in Section 4.4.

Table 5-5(a) displays the deterioration equation parameters, Table 5-5(b) the parameters obtained for the maintenance model, and Table 5-5(c) the parameters for the measurement model. As can be seen from Table 5-5(c), the units of the latent variable were fixed by setting the λ associated with the slope variance indicator to be equal to 1. As discussed in Section 4.2, this is a necessary normalization for setting the scale of the model. The latent variable is therefore measured in the units of the logarithm of the slope variance. It may be recalled that the MODEL0 and MODEL1 parameters were normalized to the same units to facilitate comparison between the models.

The fit of the model to data is indicated by the squared multiple correlation of the structural and the measurement equations. For the structural equation, the dependent variable is latent condition, and the squared multiple correlation measures the fraction of the variance of latent performance that is accounted for by the explanatory variables. It is the analog of the $R^2$ statistic, and is calculated in the usual manner by dividing the "regression sum of squares" of the structural equation by the "total sum of squares" which is derived from the regression sum of squares and the estimated error variance of the deterioration equation. For the measurement equations, the dependent variables are the extents of observed damage, and the squared multiple correlation measures the fraction of the variance of the observed variable that is accounted for by the latent variables for which the observed variable is an indicator.
### Dependent Variable: Latent Condition (Ride Quality)

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Estimate</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Maintenance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sand Seal Coat Area</td>
<td>-0.109</td>
<td>-2.75</td>
</tr>
<tr>
<td>Crack Filling Amount</td>
<td>-0.218</td>
<td>-2.86</td>
</tr>
<tr>
<td>Chip Seal Coat Area</td>
<td>-0.032</td>
<td>-0.51</td>
</tr>
<tr>
<td><strong>Traffic and Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Av. Daily Traffic</td>
<td>0.104</td>
<td>1.68</td>
</tr>
<tr>
<td>Percentage Trucks</td>
<td>2.130</td>
<td>7.49</td>
</tr>
<tr>
<td>Age</td>
<td>0.985</td>
<td>11.60</td>
</tr>
<tr>
<td>Age2</td>
<td>-0.093</td>
<td>-7.98</td>
</tr>
<tr>
<td><strong>Environmental</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Av. Annual Precipitation</td>
<td>-0.040</td>
<td>-0.43</td>
</tr>
<tr>
<td>Av. Annual Min. Temp.</td>
<td>-0.023</td>
<td>-0.38</td>
</tr>
<tr>
<td>Av. Annual Freeze/Thaw Cycles</td>
<td>-0.080</td>
<td>-1.05</td>
</tr>
<tr>
<td>Av. No. of Wet days</td>
<td>0.189</td>
<td>2.40</td>
</tr>
<tr>
<td><strong>Dummies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District 1 Dummy</td>
<td>-0.253</td>
<td>-1.74</td>
</tr>
<tr>
<td>District 2 Dummy</td>
<td>0.446</td>
<td>2.11</td>
</tr>
<tr>
<td>Road Surface Dummy</td>
<td>1.205</td>
<td>16.37</td>
</tr>
</tbody>
</table>

**No. of observations:** 3837  
**R²:** 0.63

Table 5-5(a): Deterioration Model Parameters: MODEL2
### Table 5-5(b): Maintenance Model Parameters: MODEL2

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Estimate</th>
<th>t-statistic</th>
<th>R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(1+ Slope Variance)</td>
<td>(*)&amp;1.000</td>
<td>-</td>
<td>0.40</td>
</tr>
<tr>
<td>(Cracking + Patching)^{0.5}</td>
<td>1.142</td>
<td>20.99</td>
<td>0.41</td>
</tr>
<tr>
<td>(Rut Depth)^2</td>
<td>0.323</td>
<td>7.17</td>
<td>0.02</td>
</tr>
</tbody>
</table>

(*) Required Normalization

### Table 5-5(c): Measurement Model Parameters: MODEL2

<table>
<thead>
<tr>
<th>Performance Variable</th>
<th>PSI</th>
<th>LATENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurement Variable</td>
<td>(C+P)^{a4}</td>
<td>RD^2</td>
</tr>
<tr>
<td>Coefficients</td>
<td>0.030</td>
<td>1.308</td>
</tr>
<tr>
<td>Normalized Coefficients</td>
<td>0.016</td>
<td>0.722</td>
</tr>
</tbody>
</table>

### Table 5-5(d): Comparison of PSI and Extracted Latent Variable
From Table 5-5(c), the squared multiple correlations of the individual measurement equations (presented in the table as $R^2$) have the values 0.41 for $(C + P)^{0.5}$, 0.02 for $RD^2$, and 0.39 for $\log_{10}(1 + SV)$. Since the squared multiple correlation is a measure of the variance of the indicator that is explained by the latent variable, a higher value of the $R^2$ for a particular measurement equation implies that the associated indicator is a better measure of the latent variable than an indicator with a lower value of $R^2$. From the measurement equations of MODEL2, it can be seen that the latent variable is better characterized by the slope variance and cracking rather than by the rut depth. This latent variable can therefore be thought of as a roughness related latent variable.

The relationship between the latent variable and its indicators can also be examined by constructing a "latent performance index" formula as a weighted sum of its associated indicators in the same manner as the PSI. In order to do this, it is first necessary to extract the value of the latent variable, so that it can be be regressed against the indicators to obtain the formula. The regression method described in Section 4.2 is used to obtain the value of the latent variable. The following deterioration formula is obtained:

$$\hat{S} = 0.08(C + P)^{0.5} + 1.15RD^2 + 2.79\log_{10}(1 + SV) \quad (5 - 4)$$

Table 5-5(d) lists the weights obtained for each damage measurement in the construction of the latent variable models and the PSI. From Table 5-5(d), the fundamental difference between the PSI and the latent variable estimated by MODEL2 is the importance given to the "load related" damage component, the rut depth. In the PSI equation, the ratio of the rut depth parameter to the (C+P) parameter is 46.0, while the ratio is 14.4 in equation (5-12). Similarly, the ratio of the slope variance parameter to the cracking and patching parameter is 0.016 in the PSI equation compared to 0.035 in equation (5-12), showing a greater importance given to the
roughness related measurements in the latent variable model. This is also indicated by the ratio between the rut depth parameter and the slope variance parameter, which decreased from 0.722 in the PSI equation to 0.40 for the latent variable model. The latent variable index therefore gives a greater weight to the roughness related indicators than the PSI.

From Table 5-5(a), it can be seen that the same measurements used in the latent variable model rather than in the calculation of the PSI substantially improve the fit to the data, as measured by the squared multiple correlation of the structural equation. The fit has now increased from 0.27 (in Table 5-4(a)) to 0.64. By removing the inefficiencies associated with the previously estimated PSI formula, and by giving greater importance to the measurements that directly affect ride quality, the latent variable model significantly increases the goodness-of-fit.

The formulation of MODEL2 is the direct extension of the PSI model, MODEL1. MODEL2 was specified by substituting the latent variable S for (-PSI) in equations (5-2a), (5-2b), (5-2c) and (5-2d). The latent condition S is therefore an independent variable in the maintenance equations. The assumption in this case is that the highway agencies perform maintenance by reacting to the latent condition of the pavement rather than to the actual measurements of damage. Such an assumption may characterize the long term behavior of an agency where the authorities may obtain a good perception of the latent pavement condition over time. In the shorter term, it is likely that the highway agencies may base maintenance decisions on observed damage, rather than on latent condition. The structural equation for such a model can be specified as follows:

\[ S = \beta_1A_1 + \beta_2A_2 + \beta_3A_3 + \beta_4X_1 + \eta_1 \]  

(5 - 5a)

\[ A_1 = \gamma_1f_2 + \Gamma_1X_2 + \eta_2 \]  

(5 - 5b)

\[ A_2 = \gamma_2f_2 + \Gamma_2X_2 + \eta_3 \]  

(5 - 5c)
\[ A_3 = \gamma_1 l_2 + \Gamma_3 X_2 + \eta_4 \]  

(5 - 5d)

where \( l_2 \) is a (3x1) vector of the following measurements:

1) \( \log(1 + \text{Slope Variance}) \)
2) \( (\text{Cracking} + \text{Patching})^{0.5} \)
3) \( (\text{Rut Depth})^2 \)

In addition to the explanatory variables for the maintenance equation specified in MODEL 2, equations 5-5(b), 5-5(c) and 5-5(d) specify the highway agency as reacting to three observed indicators of damage: the log of slope variance, the square root of cracking, and the rut depth.

The measurement equations remain the same as those specified in equation (5-3). The results obtained from the specification are presented in Tables 5-6(a), 5-6(b) and 5-6(c). This model is referred to as MODEL2A.

A comparison of Tables 5-5 and 5-6 shows no significant difference in the parameter estimates of the deterioration equation. No major change in the deterioration behavior of the pavement takes place due to the change in specification. The most important difference in the parameter estimates is that the traffic variable once again begins to have an incorrect sign. The reason for this is that the specification of MODEL2A does not retain any more the simultaneous equation structure of MODEL1 and of MODEL2. Since, by assumption, the error term of the deterioration equation is uncorrelated with the error terms of the maintenance equations, the maintenance model is independent of the deterioration model. However, the specification can be used to study the maintenance behavior of the agencies. From Table 5-6(b), it can be seen that the amount of maintenance performed increases for all activities as the average daily traffic increases and as the percentage of truck traffic increases. However, the amount of maintenance
Dependent Variable: Latent Condition (Ride Quality)

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Estimate</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Maintenance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sand Seal Coat Area</td>
<td>-0.090</td>
<td>-7.70</td>
</tr>
<tr>
<td>Crack Filling Amount</td>
<td>-0.213</td>
<td>-1.72</td>
</tr>
<tr>
<td>Chip Seal Coat Area</td>
<td>-0.123</td>
<td>-7.17</td>
</tr>
<tr>
<td><strong>Traffic and Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Av. Daily Traffic</td>
<td>-0.117</td>
<td>-1.59</td>
</tr>
<tr>
<td>Percentage Trucks</td>
<td>1.838</td>
<td>6.92</td>
</tr>
<tr>
<td>Age</td>
<td>0.930</td>
<td>10.50</td>
</tr>
<tr>
<td>Age2</td>
<td>-0.089</td>
<td>-7.39</td>
</tr>
<tr>
<td><strong>Environmental</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Av. Annual Precipitation</td>
<td>-0.041</td>
<td>-0.81</td>
</tr>
<tr>
<td>Av. Annual Min. Temp.</td>
<td>-0.120</td>
<td>-2.04</td>
</tr>
<tr>
<td>Av. Annual Freeze/Thaw Cycles</td>
<td>-0.088</td>
<td>-1.42</td>
</tr>
<tr>
<td>Av. No. of Wet days</td>
<td>0.330</td>
<td>4.51</td>
</tr>
<tr>
<td><strong>Dummies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District 1 Dummy</td>
<td>-0.122</td>
<td>-1.37</td>
</tr>
<tr>
<td>District 2 Dummy</td>
<td>0.294</td>
<td>3.80</td>
</tr>
<tr>
<td>Road Surface Dummy</td>
<td>1.452</td>
<td>20.71</td>
</tr>
</tbody>
</table>

No. of observations: 3837  
$R^2$: 0.67

Table 5-6(a): Deterioration Model Parameters: MODEL2A
**Dependent Variable: Maintenance Activity**

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Sand Seal Coat Estimate/t-statistic</th>
<th>Chip Seal Coat Estimate/t-statistic</th>
<th>Crack Filling Estimate/t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Av. Daily Traffic</td>
<td>0.085/2.47</td>
<td>0.030/1.16</td>
<td>0.098/8.27</td>
</tr>
<tr>
<td>Percentage Trucks</td>
<td>2.615/7.52</td>
<td>0.722/3.60</td>
<td>-0.104/-1.36</td>
</tr>
<tr>
<td>Road Surface Dummy</td>
<td>-0.909/-12.03</td>
<td>0.567/10.35</td>
<td>-0.162/-8.67</td>
</tr>
<tr>
<td>log(1 + Slope Variance)</td>
<td>-0.059/-4.05</td>
<td>-0.017/-1.54</td>
<td>0.011/3.32</td>
</tr>
<tr>
<td>(C + P)^0.5</td>
<td>-0.011/-1.22</td>
<td>-0.089/-10.07</td>
<td>0.021/6.26</td>
</tr>
<tr>
<td>(RD)^2</td>
<td>0.002/0.22</td>
<td>0.038/4.64</td>
<td>-0.001/-0.24</td>
</tr>
</tbody>
</table>

| R^2                          | 0.13                               | 0.10                               | 0.05                             |

Table 5-6(b): Maintenance Model Parameters: MODEL2A

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Estimate</th>
<th>t-statistic</th>
<th>R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(1+ Slope Variance)</td>
<td>(*)1.000</td>
<td>-</td>
<td>0.42</td>
</tr>
<tr>
<td>(Cracking + Patching)^0.5</td>
<td>1.094</td>
<td>24.53</td>
<td>0.39</td>
</tr>
<tr>
<td>(Rut Depth)^2</td>
<td>0.339</td>
<td>7.75</td>
<td>0.02</td>
</tr>
</tbody>
</table>

(* Required Normalization

Table 5-6(c): Measurement Model Parameters: MODEL2A
performed varies substantially with changes in observed damage. The amount of crack filling performed increases as the roughness increases, the extent of cracking increases and as the rut depths increase. The other two maintenance activities, however, decrease as the roughness increases, and as the extent of cracking and patching increases. This is consistent with the role played by different maintenance activities in correcting damage as reported by Humplick[1984]. As mentioned before, according to current highway practice, seal coating operations cannot be performed without crack filling for cracks that have more than hairline width, so it is reasonable to expect that the amount of seal coating decreases as the cracking increases. Also, seal coating is a patching operation for water proofing the pavement, and contributes to increased roughness of the pavement. This is reflected in the negative sign of the parameter associated with slope variance for the two seal coat activities.

5.5 A Latent Variable Model with Two Latent Variables

The models presented in the previous section, MODEL2 and MODEL2A are specifications involving a single latent variable. As mentioned before, this latent variable relates to the roughness or the ride quality of the pavement. Experiments with other measurements in the data indicated the presence of a second latent variable that is related to load bearing capacity of the pavement, and whose indicators are alligator cracking, and the width (severity) of cracks. An example of experiments of this kind are described in Section 6.1. Since alligator cracking is related to structural damage arising from traffic loads, this latent variable may be termed "structure related" or "load related" and may possibly include the rut depth as an indicator as well. The two latent variable model is specified in the same fashion as MODEL2, except that there is now a second deterioration equation that predicts the deterioration of the load related latent variable.
This model, called MODEL3, can be specified as follows:

The structural equations are:

\[ S_1 = \beta_1 A_1 + \beta_2 A_2 + \beta_3 A_3 + \beta_4 X_1 + \eta_1 \]  
\[ (5 - 6a) \]

\[ S_2 = \beta_5 X_3 + \eta_2 \]  
\[ (5 - 6b) \]

\[ A_1 = \gamma_1 S_1 + \Gamma_1 X_2 + \eta_3 \]  
\[ (5 - 6c) \]

\[ A_2 = \gamma_2 S_1 + \Gamma_2 X_2 + \eta_4 \]  
\[ (5 - 6d) \]

\[ A_3 = \gamma_3 S_1 + \Gamma_3 X_2 + \eta_5 \]  
\[ (5 - 6e) \]

where \( S_1 \) refers to the roughness related latent variable, \( S_2 \) refers to the load related latent variable, the vectors \( X_1 \) and \( X_2 \) are explanatory variable vectors described in equation (5-2). The vector \( X_3 \) is a vector of explanatory variables relating to the second latent variables \( S_2 \). Since \( S_2 \) reflects the structural strength of the pavement, it is unlikely to depend on surface routine maintenance activities or on environment. The specification of equation (5-6b) therefore includes only the following:

a) Average Daily Traffic

b) Percentage Trucks

The measurement equations are as follows:

\[ (C + P)^{0.5} = \lambda_1 S_1 + \varepsilon_1 \]  
\[ (5 - 7a) \]

\[ RD^2 = \lambda_2 S_1 + \varepsilon_2 \]  
\[ (5 - 7b) \]
\log_{10}(1 + SV) = \lambda_3 S_1 + \varepsilon_3 \hspace{1cm} (5-7c)

Al\text{-cracking extent} = \lambda_4 S_2 + \varepsilon_4 \hspace{1cm} (5-7d)

Al\text{-Width} = \lambda_5 S_2 + \varepsilon_5 \hspace{1cm} (5-7e)

Lin\text{-Width} = \lambda_6 S_2 + \varepsilon_6 \hspace{1cm} (5-7f)

rut depth = \lambda_7 S_2 + \varepsilon_7 \hspace{1cm} (5-7g)

Examination of equations (5-6) and (5-7) reveals that the model system can be broken up into 2 parts:

a) PART 1: Equations (5-6a), (5-6c) and (5-7a) through (5-7c). These equations specify the deterioration model for the roughness related latent variable, as formulated in MODEL2. Equation (5-7a), (5-7b) and (5-7c) are the three measurement equations for the roughness related latent variable, as specified in MODEL2.

b) PART 2: Equations (5-6b) and (5-7d), (5-7e), (5-7f) and (5-7g). Equation (5-6b) specifies the deterioration model for the load related latent variable. Equations (5-7d), (5-7e), (5-7f) and (5-7g) are the measurement equations for this latent variable, with the extent of alligator cracking, the average width of alligator and linear cracks and the average rut depths being the indicators.

In the absence of correlation between the error terms of the two parts described above, the model parts are independent, so part 2 can be separately estimated. Part 1 has exactly the same specification as MODEL2, and so produces the same parameter estimates. The estimation of the parameters of part 2 can be done separately using equations (5-6b) and (5-7d), (5-7e), (5-7f) and (5-7g). The results of this model are presented in Tables 5-7(a) and 5-7(b) as MODEL3.
### Table 5-7(a): Deterioration Model Parameters: MODEL3

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Estimate</th>
<th>t-statistic</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Alligator Cracking)(^{0.5})</td>
<td>(*)1.000</td>
<td>-</td>
<td>0.49</td>
</tr>
<tr>
<td>All. Cracking Width</td>
<td>5.326</td>
<td>12.33</td>
<td>0.94</td>
</tr>
<tr>
<td>Lin. Crack Width</td>
<td>0.129</td>
<td>8.55</td>
<td>0.08</td>
</tr>
<tr>
<td>Rut Depth</td>
<td>1.682</td>
<td>8.73</td>
<td>0.07</td>
</tr>
</tbody>
</table>

(*) Required Normalization

### Table 5-7(b): Measurement Model Parameters: MODEL3
The measurement model of Table 5-7(b) shows that the alligator cracking extent and the alligator widths are indeed indicators for the second latent variable, since the squared multiple correlations for the associated measurement equation are 0.65, and 0.71 respectively. However, the deterioration model has a very poor fit to the data, with a squared multiple correlation of only 0.05. This indicates that the deterioration relationship for the load related latent variable is not well specified. The solution to this problem would be to seek other explanatory variables which could serve as independent variables in the deterioration equation (5-6b). Such variables could be a more complete breakdown of the different traffic load categories, or the materials with which the sub surface layers are constructed.

The assumption of orthogonality between the two latent variables is also not very realistic. The interaction between the two latent variables might arise due to many reasons. In MODEL3, the same measurements are indicators for more than one latent variable; for example, the square of the rut depth is used as an indicator for the roughness related latent variable, \( S_1 \), while the rut depth itself is used as an indicator for \( S_2 \). This means that the degradation in the ride quality and a deterioration in the structural strength both manifest themselves in the same observed damage. These two measurements are therefore correlated. Similarly, the sub-surface processes that give rise to deterioration and produce different kinds of surface damage are likely to be correlated. This would cause a correlation in the error terms of the two deterioration equation, i.e., of equation (5-6a) and (5-6b).

From the preceding discussion, the following are the requirements for a better model specification:

a) The data gives a clear indication of more than one latent variable, and the specification of the roughness related latent variable \( S_1 \) may be substantially improved by the introduction of
correlation between the error terms \( \eta_1 \) and \( \eta_2 \) of the two deterioration equations.

b) There is a substantial overlap between the indicators of the two latent variables, and this might give rise to correlations between the measurements. This correlation can be directly captured by loading both latent variables on the same indicators.

c) A more complete specification for the structural equation of the load related latent variable \( S_2 \) is required.

With these three requirements in mind, a final model, denoted MODEL4, is presented in Tables 5-8(a), 5-8(b) and 5-8(c). Similarity in the unobserved aspects of deterioration that affect the degradation of ride quality and structural strength are captured by introducing correlations between the error terms of the deterioration equation of the two latent variables. Other correlations, for example between the \( \eta \)'s and the \( \epsilon \)'s are still assumed to be zero. In addition, both latent variables are loaded on the same indicators if overlap occurs. Finally, since \( S_2 \) relates to the structure of the pavement, it is likely to be affected by cumulative effects such as age, cumulative traffic and the number of freeze-thaw cycles. The specification of the structural model for the load related latent variable is improved by the addition of these effects. The model specification is as follows:

The structural equations are:

\[
S_1 = \beta_1^1 A_1 + \beta_2^1 A_2 + \beta_3^1 A_3 + \beta_4^1 X_1 + \eta_1 \quad (5-8a)
\]

\[
S_2 = \beta_5^2 X_3 + \eta_2 \quad (5-8b)
\]

\[
A_1 = \gamma_1 S_1 + \Gamma_1 X_2 + \eta_3 \quad (5-8c)
\]
### Dependent Variable: Latent Condition (Ride Quality)

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Estimate</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Maintenance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sand Seal Coat Area</td>
<td>-0.076</td>
<td>-5.75</td>
</tr>
<tr>
<td>Crack Filling Amount</td>
<td>-0.219</td>
<td>-3.18</td>
</tr>
<tr>
<td>Chip Seal Coat Area</td>
<td>-0.116</td>
<td>-4.33</td>
</tr>
<tr>
<td><strong>Traffic and Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Av. Daily Traffic</td>
<td>0.238</td>
<td>1.67</td>
</tr>
<tr>
<td>Percentage Trucks</td>
<td>0.027</td>
<td>0.06</td>
</tr>
<tr>
<td>Age</td>
<td>0.662</td>
<td>6.88</td>
</tr>
<tr>
<td>Age2</td>
<td>-0.059</td>
<td>-4.78</td>
</tr>
<tr>
<td><strong>Environmental</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Av. Annual Precipitation</td>
<td>-0.067</td>
<td>-0.99</td>
</tr>
<tr>
<td>Av. Annual Min. Temp.</td>
<td>0.150</td>
<td>2.57</td>
</tr>
<tr>
<td>Av. Annual Freeze/Thaw Cycles</td>
<td>-0.135</td>
<td>-2.02</td>
</tr>
<tr>
<td>Av. No. of Wet days</td>
<td>-0.078</td>
<td>-0.86</td>
</tr>
<tr>
<td><strong>Dummies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District 1 Dummy</td>
<td>-0.319</td>
<td>-1.31</td>
</tr>
<tr>
<td>District 2 Dummy</td>
<td>0.167</td>
<td>1.29</td>
</tr>
<tr>
<td>Road Surface Dummy</td>
<td>1.381</td>
<td>15.30</td>
</tr>
</tbody>
</table>

No. of observations: 3837
R²: 0.89

Table 5-8(a): Deterioration Model 1 Parameters: MODEL4
### Table 5-8(b): Deterioration Model 2 Parameters: MODEL4

**Dependent Variable: Maintenance Activity**

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Sand Seal Coat Estimate/t-statistic</th>
<th>Chip Seal Coat Estimate/t-statistic</th>
<th>Crack Filling Estimate/t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>0.049/0.76</td>
<td>-0.298/-4.81</td>
<td>0.093/6.24</td>
</tr>
<tr>
<td>Av. Daily Traffic</td>
<td>0.089/2.35</td>
<td>-0.009/-0.29</td>
<td>0.114/8.83</td>
</tr>
<tr>
<td>Percentage Trucks</td>
<td>2.569/7.63</td>
<td>0.779/3.63</td>
<td>-0.052/-0.63</td>
</tr>
<tr>
<td>Road Surface Dummy</td>
<td>-0.955/-7.50</td>
<td>0.908/7.29</td>
<td>-0.955/-7.587</td>
</tr>
</tbody>
</table>

\[ R^2 = 0.13 \]

### Table 5-8(c) Maintenance Model Parameters: MODEL4

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Estimate</th>
<th>t-statistic</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \log(1 + \text{Slope Variance}) S_i )</td>
<td>(*)1.000</td>
<td>-</td>
<td>0.25</td>
</tr>
<tr>
<td>( (\text{Cracking + Patching})^{ae} S_i )</td>
<td>0.632</td>
<td>10.84</td>
<td>0.69</td>
</tr>
<tr>
<td>( (\text{Cracking + Patching})^{ae} S_i )</td>
<td>(*)1.000</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>( (\text{Rut Depth})^{3} S_i )</td>
<td>-0.545</td>
<td>-1.31</td>
<td>0.05</td>
</tr>
<tr>
<td>( (\text{Rut Depth})^{3} S_i )</td>
<td>0.795</td>
<td>3.31</td>
<td></td>
</tr>
<tr>
<td>All. Cracking Width ( S_i )</td>
<td>0.599</td>
<td>8.45</td>
<td>0.64</td>
</tr>
<tr>
<td>Lin. Crack Width ( S_i )</td>
<td>0.269</td>
<td>7.60</td>
<td>0.11</td>
</tr>
</tbody>
</table>

(*) Required Normalization

### Table 5-8(d): Measurement Model Parameters: MODEL4

137
\[ A_2 = \gamma_2 S_1 + \Gamma_2 X_1 + \eta_4 \quad (5-8d) \]

\[ A_3 = \gamma_3 S_1 + \Gamma_3 X_1 + \eta_5 \quad (5-8e) \]

where \( S_1 \) refers to the roughness related latent variable, \( S_2 \) refers to the load related latent variable and the vectors \( X_1 \) and \( X_2 \) are explanatory variable vectors described in equation (5-2). As in MODEL3, \( X_3 \) is a vector of explanatory variables relating to the second latent variable \( S_2 \). The vector \( X_3 \) is respecified with the following variables:

1) Age
2) \( Age^2 \)
3) Cumulative Traffic (measured as ADT*AGE)
4) Cumulative Truck Traffic
5) Road Surface type
6) Freeze thaw cycles

The measurement equations for the model are follows:

\[(C + P)^{0.5} = \lambda_{11} S_1 + \lambda_{12} S_2 + \varepsilon_1 \quad (5-9a)\]

\[RD^2 = \lambda_{21} S_1 + \lambda_{22} S_2 + \varepsilon_2 \quad (5-9b)\]

\[\log_{10}(1 + SV) = \lambda_{31} S_1 + \varepsilon_3 \quad (5-9c)\]

\[Al.Width = \lambda_{42} S_2 + \varepsilon_4 \quad (5-9d)\]

\[Lin.Width = \lambda_{52} S_2 + \varepsilon_5 \quad (5-9e)\]
Results for this model are shown in Tables 5-8(a), 5-8(b), 5-8(c) and 5-8(d). Examination of Tables 5-8(a) and 5-8(b) show that the significant variables have the expected signs. The effect of traffic is not significant in either of the deterioration equations, but the effect of cumulative truck traffic is significant in the deterioration equation for structural strength. The effect of age is significant in both the deterioration equations, and the number of freeze-thaw cycles is significant in explaining load related deterioration. The parameters in Table 5-8(c) show that the effects of condition on the amount of maintenance performed (first row of Table 5-8(c)) have positive signs for sand seal coat and crack filling, but negative signs for chip seal coat. The negative sign is consistent with the findings from MODEL2A that seal coating activities are not triggered by roughness related damage. From the measurement model parameters of Model 5-8(d), it can be seen that rut depth loading parameter $\lambda_{21}$ is not significant for the ride quality variable, but $\lambda_{22}$ is significant for the structural strength latent variable. However, the dominant indicator for the structural latent variable is cracking, and this is illustrated by the high values of the coefficient of determination for the two cracking related measurements of MODEL4 (0.68 and 0.64 in Table 5-8(d)). MODEL4 substantially improves the fit of the ride quality deterioration equation with a squared multiple correlation value of 0.89. The fit of the deterioration equation for the load related latent variable has also increased from 0.05 to 0.2.

The latent variable $S_1$ estimated in MODEL4 is less similar to the PSI than the latent variable of MODEL2. This is apparent from the correlations between the extracted value of the dependent variable of the deterioration equation for each model. This value is calculated from the estimated structural equations for MODEL1, and extracted in MODEL2 and MODEL4 using the factor extraction procedure described previously. The correlation between $-\text{PSI}$ for MODEL1 and $\hat{S}$ for MODEL2 is 0.9, indicating that both these variables are modeling the same deterioration characteristic. However, the correlation between MODEL1 and $\hat{S}_1$ of
MODEL4 reduces to 0.61. This indicates that the latent variable of MODEL4 is including some other characteristics that influence roughness that are not modeled by the PSI or the single latent variable of MODEL2.

An indication of this becomes clearer on examining some statistics of the latent variables extracted from MODEL2 and MODEL4, and comparing these statistics with the extracted PSI from MODEL1. The PSI is extracted from the reduced form equations obtained by substituting the maintenance equations into the deterioration model. Table 5-9(a) shows summary statistics for the roughness related latent variables from MODEL2 and MODEL4 and the PSI. All the performance variables are in units of log(slope variance).

From Table 5-9(a), it is seen that the fitted value of PSI has a higher variance than either of the latent variables. The variances of the extracted values of the two latent variables are very similar. However, the correlation between the two latent variables is only 0.63. This is because the latent variable from MODEL2 is very highly correlated with the PSI, and the roughness related latent variable is somewhat different from the PSI. The coefficient of kurtosis is close to 3.0 for all the extracted performance variables of Table 5-9(a), indicating that they are normal or near normally distributed.

Tables 5-9(b) and 5-9(c) demonstrate once again the difference between the extracted latent variable and the fitted PSI. In these tables, the pavements have been divided into four categories: a) poor b) fair c) good d) excellent. Each of these categories includes 25% of the data, with "poor" being the 25% of the pavements in the worst condition, and "excellent" being the 25% of the pavements in the best condition. The tables show the number of pavements whose condition falls into each category, as measured by the latent variables and the PSI. If the latent variables are highly correlated, or are correlated with the PSI, then a large number of obser-
<table>
<thead>
<tr>
<th>Performance Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSI (MODEL1)</td>
<td>-5.60</td>
<td>2.75</td>
<td>0.00</td>
<td>2.79</td>
</tr>
<tr>
<td>S (MODEL2)</td>
<td>-3.50</td>
<td>2.77</td>
<td>0.00</td>
<td>1.39</td>
</tr>
<tr>
<td>S₁ (MODEL4)</td>
<td>-3.78</td>
<td>4.14</td>
<td>0.00</td>
<td>1.46</td>
</tr>
</tbody>
</table>

Table 5-9(a): Statistics of Extracted Performance Variables

**RIDE QUALITY (MODEL2)**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Excellent</th>
<th>Good</th>
<th>Fair</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSI</td>
<td>815</td>
<td>137</td>
<td>12</td>
<td>00</td>
</tr>
<tr>
<td>Good</td>
<td>139</td>
<td>555</td>
<td>215</td>
<td>45</td>
</tr>
<tr>
<td>Fair</td>
<td>05</td>
<td>213</td>
<td>466</td>
<td>273</td>
</tr>
<tr>
<td>Poor</td>
<td>00</td>
<td>54</td>
<td>266</td>
<td>642</td>
</tr>
</tbody>
</table>

Table 5-9(b): Condition Prediction by MODEL2 and PSI

**RIDE QUALITY (MODEL4)**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Excellent</th>
<th>Good</th>
<th>Fair</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSI</td>
<td>665</td>
<td>208</td>
<td>84</td>
<td>07</td>
</tr>
<tr>
<td>Good</td>
<td>251</td>
<td>184</td>
<td>178</td>
<td>341</td>
</tr>
<tr>
<td>Fair</td>
<td>40</td>
<td>210</td>
<td>398</td>
<td>309</td>
</tr>
<tr>
<td>Poor</td>
<td>01</td>
<td>357</td>
<td>300</td>
<td>303</td>
</tr>
</tbody>
</table>

Table 5-9(c): Condition Prediction by MODEL4 and PSI
vations will be predicted to be in the same category by both performance measures.

Since each category represents a quarter of the population, the total number of observations in each category is about 960. Table 5-9(b), which compares the condition measured by the PSI with the condition measured from MODEL2 shows the very high correlation between the two performance measures. 815 pavements fall in the uppermost quartile in terms of both performance variables, indicating that over 85% of the "excellent" pavements are categorized similarly by both measures. There is no large bias apparent in the number of observations that are classified differently; 139 observations classified by the latent variable as being "excellent" are classified by the PSI as being "good"; the reverse is true for 137 pavements. Some of the differences in categorization may occur at the boundary between two classes, reflecting no real difference in predicted condition. The categorizations from the two models are closer at either end of the scale; i.e for pavements that are "excellent" or "poor" than in the middle, where about 60% of the pavements in the "good" category and about 50% in the "fair" category are classified similarly. The difference in the classifications stems from the difference in the characteristics measured by the latent variable and the PSI, since the latent variable places a greater emphasis on cracking and slope variance, and less emphasis on the rut depth than the PSI. However, in general, pavements in good condition are smooth and rut-free, while pavements in poor condition are rough and rut-ridden, accounting for the greater degree of correspondence at either ends. In the middle, relatively smooth but somewhat rutted pavements or pavements with the reverse characteristics may be encountered. This might be the reason for the poorer correspondence in the "good" and "fair" regions.

Table 5-9(c) that presents the same information as Table 5-9(b) for the PSI and the latent variable $S_1$ from MODEL4. As can be seen from this table, the correspondence is poorer. The correspondence is the closest for pavements in the "excellent" category, with about 70% of the
observations being identically classified by both measures, but for other categories, the percentage of correspondence is substantially lower. For example, in the "poor" category, only 24% of the pavements are classified identically. Of the 959 pavements that are "poor" as measured by the latent variable, 309 as classified as "fair" and 341 as "good" by the PSI.

It is clear from Table 5-9(c) that the latent variable \( S_i \) of MODEL4 represents pavement characteristics other than those represented by the PSI. Some of these characteristics are unobserved correlations between the cracking process and the roughness propagation process; these are incorporated in MODEL4 through the correlation between the error terms. The second major difference between the PSI and \( S_i \) is the importance of rut depth. The latent variable model of MODEL4 gives less importance to the rut depth than the PSI, as can be seen from Table 5-8(d), where the rut depth term has an unexpected sign and is not significant. These differences might account for the differences in classification between the PSI and the latent variable of MODEL4. Clearly, it is not possible to state that one measure is "better" than the other, since they are not measuring the same characteristics; however, the performance characteristic from MODEL4 fits the data much better than the PSI model (with a squared multiple correlation of 0.89); also, its distribution has a smaller variance than that of the PSI. The only point that can be explicitly made is that maintenance decisions based on the latent variable of MODEL4 may have different implications from maintenance decisions made using the PSI. This should be considered in evaluating the output from traditional strategy selection models, many of which are based on the PSI.

In conclusion, MODEL4 is a specification for a predictive model for ride quality and load related characteristics of a highway pavements that is substantially different from the first model described in this chapter, i.e. the "naive" deterioration model MODEL0 based on the PSI. MODEL4 incorporates the following changes:
a) Removal of the reliance of the deterioration model on judgemental formulae or selection of a small number of indicators.

b) Incorporation of the simultaneous effect of maintenance and deterioration.

c) Incorporation of the interaction between ride quality and load bearing capacity of a pavement.

Incorporation of the above characteristics has produced a realistic model with appropriate signs of the parameter estimates and an improved fit to data.

5.6 Conclusion

This chapter has demonstrated the use of a method for modeling pavement deterioration that improves on traditional methods through the use of a rigorous statistical procedure on field data. The deterioration model is estimated using a latent dependent variable whose value is simultaneously estimated from a set of indicators, which are measures of damage on the pavement. The latent variable modeling approach has the flexibility to combine many different types of observed data into the specification of a deterioration model, and provides a guideline for differentiating between damage measurements that contributes to the specification, and those that do not. This aspect will be discussed in Chapter 6.

Apart from the latent variable approach, this chapter also implemented the deterioration prediction model system as a system of simultaneous equations to account for the interacting processes of deterioration and the agency’s maintenance behavior. This specification solved some of the problems with unreasonable signs for parameter estimates that have been reported in the literature on analysis performed with the same data set.

The results from this chapter indicate that the latent variable approach has the potential to estimate realistic infrastructure deterioration models, which can then be used for better planning
of facility maintenance strategies. The manner in which this can be done is described in the following chapter.
CHAPTER 6
APPLICATIONS OF LATENT VARIABLE DETERiorATION MODELS

Chapter 5 presented a case study that implemented the latent variable approach on pavement damage data and compared the performance measures obtained from the latent variable models with the PSI. The results indicated that the latent variable approach has the potential to obtain performance measures that fit the data better than the PSI. However, in addition to the theoretical considerations, there is also the requirement that the model should be applicable in the field, and that the information obtained from the model should be useful to the authorities involved in the management of infrastructure facilities. The two most important applications of a deterioration prediction model are the following:

a) The model should be able to provide some indication of the data that should be collected by the agency for performance evaluation, and maintenance planning. Since data collection is an expensive procedure, it would be useful if the model provided information on the extent to which a particular measurement or a group of measurements contribute to the prediction of performance. This information can then be used to select the types of data that need to be collected, and subsequently to plan a data collection procedure.

b) The model should be able to predict the impacts of changes in the explanatory variables on the performance of the pavement. It should be possible to calculate measures such as the elasticity of performance with respect to important explanatory variables such as traffic or age, so that the sensitivity of performance to these variables can be determined. This information is useful for forecasting the performance of the pavement under a variety of different future scenarios. These forecasts are useful for planning future maintenance to be performed on the pavement. The model should also be able to provide some probabilistic measure of the accuracy
of predicted performance, such as the distribution (or at least the moments of the distribution) of the performance measure. This will assist in the development of probabilistic models for the selection of maintenance strategies. These probabilistic models are much more realistic than deterministic models based only on expected values.

In this chapter, the manner in which the latent variable model can be used to satisfy the above requirements are described. The first part of the chapter presents a methodology for distinguishing between different indicators for the latent variable. The second part describes the manner in which the latent variable model can be used to predict the future condition of the pavement.

6.1 Effect of Additional Measurements:

In this section the effect of adding additional measurement equations to an existing latent variable specification is investigated. The base model to which measurements are added is the PSI type latent variable specification discussed in Section 5.4 (MODEL2), with results presented in Tables 5-5(a), 5-5(b) and 5-5(c). The model specification, from Section 5.4, is repeated below:

Structural Model:

\[
S = \beta_1 A_1 + \beta_2 A_2 + \beta_3 A_3 + \beta_4 X_1 + \eta_1 \tag{6-1a}
\]

\[
A_1 = \gamma_1 S + \Gamma_1 X_2 \tag{6-1a}
\]

\[
A_2 = \gamma_2 S + \Gamma_2 X_2 \tag{6-1b}
\]

\[
A_3 = \gamma_3 S + \Gamma_3 X_2 \tag{6-1c}
\]
where $S$ is the latent performance variable, $A_1, A_2, A_3$ are the extents of sand sealing, crack filling and chip sealing respectively, $X_1$ is a vector of explanatory variables that affect performance, and $X_2$ is a vector of explanatory variables that affect the amount of maintenance performed. This specification is described in detail in Section 5.4.

Measurement Model:

$$(C + P)^{0.5} = \lambda_1 S + \epsilon_1 \quad (6 - 2a)$$

$$RD^2 = \lambda_2 S + \epsilon_2 \quad (6 - 2b)$$

$$\log_{10}(1 + SV) = \lambda_3 S + \epsilon_3 \quad (6 - 2c)$$

The latent variable obtained from this specification was seen to be highly correlated with the PSI, but was an even more direct measure of roughness than the PSI, because it gave more weight to cracking and patching and less weight to the rut depth than the PSI.

In order to investigate the effect of additional measurements, the following measurement equations were added to the model described above:

$$\text{al.sev.} = \lambda_4 S + \epsilon_4 \quad (6 - 2d)$$

$$\text{lin.sev.} = \lambda_5 S + \epsilon_5 \quad (6 - 2e)$$

where \text{al.sev.} and \text{lin.sev.} represent the severities of alligator and linear cracking as measured by the widths of the cracks in inches. Usually, wide cracks or high severity cracking is an indication of sub-surface structural damage on the pavement, so it is anticipated that these
measurements will not contribute to the prediction of a roughness or surface quality related latent variable. The following criteria can be used to test the contribution of a measurement to an associated latent variable:

a) Squared multiple correlation of the associated measurement equation. If the new measurement equation has a high squared multiple correlation, then it is a strong indicator for the latent variable.

b) Squared multiple correlation of the structural equation. If the introduction of new measurements increases the squared multiple correlation of the structural equation, then the measurement is contributing positively to the prediction of the performance variable and should be included.

The application of the criteria mentioned above is not always straightforward. To see this, consider the model formulated in equations (6-1) and (6-2). The results from this model, called MODEL2X are tabulated in Tables 6-1(a), 6-1(b) and 6-1(c). From Table 5-5(a), the squared multiple correlation for the deterioration equation of MODEL2 is 0.63. By comparison, the squared multiple correlation for the deterioration equation of MODEL2X (from Table 6-1(a)) has decreased to 0.34. By the second test described above, this is an indication that the measurements do not contribute positively to the prediction of the performance variable. However, an examination of the squared multiple correlations of the measurement equations from Table 6-1(c) shows that the measurement equation for alligator cracking severity has a squared multiple correlation of 0.55, indicating that the measurement is useful in explaining the effect of the associated latent variable S. From Table 6-1(c), the squared multiple correlation of the five measurement equations are seen to be 0.75, 0.02, 0.24, 0.55 and 0.09. The corresponding values for the three measurement equations 6-2(a), 6-2(b) and 6-2(c) from MODEL2 are 0.42,
Dependent Variable: Latent Condition (Ride Quality)

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Estimate</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Maintenance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sand Seal Coat Area</td>
<td>-0.026</td>
<td>-0.4</td>
</tr>
<tr>
<td>Crack Filling Amount</td>
<td>-0.534</td>
<td>-3.15</td>
</tr>
<tr>
<td>Chip Seal Coat Area</td>
<td>0.146</td>
<td>1.59</td>
</tr>
<tr>
<td><strong>Traffic and Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Av. Daily Traffic</td>
<td>0.068</td>
<td>1.43</td>
</tr>
<tr>
<td>Percentage Trucks</td>
<td>2.045</td>
<td>6.13</td>
</tr>
<tr>
<td>Age</td>
<td>0.623</td>
<td>7.32</td>
</tr>
<tr>
<td>Age2</td>
<td>-0.060</td>
<td>-5.04</td>
</tr>
<tr>
<td><strong>Environmental</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Av. Annual Precipitation</td>
<td>0.087</td>
<td>0.61</td>
</tr>
<tr>
<td>Av. Annual Min. Temp.</td>
<td>0.014</td>
<td>0.04</td>
</tr>
<tr>
<td>Av. Annual Freeze/Thaw Cycles</td>
<td>0.076</td>
<td>0.50</td>
</tr>
<tr>
<td>Av. No. of Wet days</td>
<td>0.121</td>
<td>0.24</td>
</tr>
<tr>
<td><strong>Dummies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District 1 Dummy</td>
<td>-0.091</td>
<td>-0.21</td>
</tr>
<tr>
<td>District 2 Dummy</td>
<td>0.653</td>
<td>2.14</td>
</tr>
<tr>
<td>Road Surface Dummy</td>
<td>0.483</td>
<td>4.47</td>
</tr>
</tbody>
</table>

No. of observations: 3837
R²: 0.33

Table 6-1(a): Deterioration Equation: MODEL2X
Dependent Variable: Maintenance Activity

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Sand Seal Coat Estimate/t-statistic</th>
<th>Chip Seal Coat Estimate/t-statistic</th>
<th>Crack Filling Estimate/t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>0.057/0.32</td>
<td>-0.460/-3.01</td>
<td>0.167/4.79</td>
</tr>
<tr>
<td>Av. Daily Traffic</td>
<td>0.106/2.24</td>
<td>0.015/0.44</td>
<td>0.108/6.96</td>
</tr>
<tr>
<td>Percentage Trucks</td>
<td>2.451/5.18</td>
<td>1.584/4.46</td>
<td>-0.348/-2.80</td>
</tr>
<tr>
<td>Road Surface Dummy</td>
<td>-1.090/-4.78</td>
<td>0.806/4.98</td>
<td>-0.260/-4.64</td>
</tr>
</tbody>
</table>

R²  

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Estimate</th>
<th>t-statistic</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(1+ Slope Variance)</td>
<td>(*)1.000</td>
<td>-</td>
<td>0.25</td>
</tr>
<tr>
<td>(Cracking + Patching) ⁰⁵</td>
<td>1.974</td>
<td>9.7</td>
<td>0.75</td>
</tr>
<tr>
<td>(Rut Depth) ²</td>
<td>0.442</td>
<td>9.36</td>
<td>0.02</td>
</tr>
<tr>
<td>Al. Cracking Width</td>
<td>0.722</td>
<td>17.19</td>
<td>0.55</td>
</tr>
<tr>
<td>Lin. Cracking Width</td>
<td>0.323</td>
<td>15.23</td>
<td>0.09</td>
</tr>
</tbody>
</table>

(*) Required Normalization

Table 6-1(b): Maintenance Model Parameters: MODEL2X

Table 6-1(c): Measurement Model Parameters: MODEL2X
0.02 and 0.40 from Table 5-5(c). In MODEL2, the roughness measurement equation has as
good a fit as the cracking measurement equation; however, in MODEL2X, the cracking
measurement equations have a much higher fit than the roughness measurements. This means
that the latent variable in the deterioration equation of MODEL2X is not a roughness related
variable any more, but a cracking related latent variable with characteristics different from the
latent variable of MODEL2. This is the reason for the decrease in the fit of the structural
equation as the result of the addition of the two measurements. The addition of alligator and
linear cracking severities to the measurement model system caused a change in the latent
variable being modeled; therefore, while the overall fit of the structural equation was not as
good as that of MODEL2, the fit of some measurement equations (including one of the new
ones) actually improved over MODEL2.

A "performance formula" in terms of the three measurements that make up the PSI was cal-
culated for MODEL2X, by regressing the extracted latent variable (using the techniques
described in Section 5.4) against the three measurements. The results are shown in Table 6-2(a),
where the weights obtained for each measurement from the regression are compared with the
weights used in the calculation of the PSI, and also with the corresponding weights used in the
calculation of the latent variable from MODEL2, which is highly correlated with the PSI. As
can be seen from this table, the weights attached to the measurements 0.12, 1.47 and 0.87 for
cracking, rut depth and log(slope variance) respectively. The corresponding figures for the
PSI are 0.03, 1.38 and 1.91, and for the latent variable of MODEL2, presented for comparison
in Table 6-2(b) are 0.08, 1.15 and 2.79. Compared to the latent variable of MODEL2 or the
PSI, it can be seen that MODEL2X assigns a much higher weight to the components of cracking.
As a result, the latent variable of MODEL2X is different from the PSI or the "roughness related"
latent variable of MODEL2. This can also be seen by the correlation between the FSI and the
<table>
<thead>
<tr>
<th>Performance Variable</th>
<th>PSI</th>
<th>LATENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurement Variable</td>
<td>(C+P)^a^b</td>
<td>RD^2</td>
</tr>
<tr>
<td>Coefficients</td>
<td>0.030</td>
<td>1.308</td>
</tr>
<tr>
<td>Normalized Coefficients</td>
<td>0.016</td>
<td>0.722</td>
</tr>
</tbody>
</table>

*Table 6-2(a): Comparison of PSI and Extracted Latent Variable (PSI and MODEL2X)*

<table>
<thead>
<tr>
<th>Performance Variable</th>
<th>PSI</th>
<th>LATENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurement Variable</td>
<td>(C+P)^a^b</td>
<td>RD^2</td>
</tr>
<tr>
<td>Coefficients</td>
<td>0.030</td>
<td>1.308</td>
</tr>
<tr>
<td>Normalized Coefficients</td>
<td>0.016</td>
<td>0.722</td>
</tr>
</tbody>
</table>

*Table 6-2(b): Comparison of PSI and Extracted Latent Variable (PSI and MODEL2)*
extracted values of the two latent variables. The correlation between the PSI and the extracted roughness related latent variable of MODEL2 is 0.9; between the PSI and the extracted cracking related variable of MODEL2X, the correlation drops to 0.61.

The question that is being investigated is whether the two additional measurements, i.e. of alligator cracking width and linear cracking width are useful for the specification of latent variable deterioration models, and whether data on these measurements should be collected in the future. From the discussion of the preceding paragraph, the answer to this question is not simple. If the object of interest was only to model the progression of ride quality, then the two severity measurements are not useful and should be dropped. However, if the modeling of overall deterioration is of interest, then the two severity measures should be retained, but the specification should include an additional latent variable that is cracking related. This latent variable should have as its indicators the measurements on the severity of alligator and longitudinal cracking. This was the idea behind the specification the two variable latent model (MODEL4) in Chapter 5. In this case, as discussed in Chapter 5, the addition of the second latent variable actually improved the fit of the ride quality deterioration model, by increasing the squared multiple correlation from 0.63 to 0.89.

The fact that latent variable of MODEL2X is very similar to the cracking related (or load related) latent variable is seen from Table 6-3, which tabulates the number of pavements in four categories: "poor", "fair", "good" and "excellent" where each category corresponds to 25% of the data. The number of pavements classified into each category on the basis of the latent variable from MODEL2X and the cracking related latent variable $S_2$ from MODEL4 is presented. As can be seen from the table, there is a good fit between the two latent variables, with 70 to 85% of the observations being classified in the same category by the two latent variables. The correlation between the two latent variable is 0.96.
<table>
<thead>
<tr>
<th>Condition</th>
<th>Excellent</th>
<th>Good</th>
<th>Fair</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>LATENT</td>
<td>Excellent</td>
<td>804</td>
<td>154</td>
<td>00</td>
</tr>
<tr>
<td>VARIABLE</td>
<td>Good</td>
<td>155</td>
<td>655</td>
<td>250</td>
</tr>
<tr>
<td>(MODEL4)</td>
<td>Fair</td>
<td>00</td>
<td>149</td>
<td>702</td>
</tr>
<tr>
<td></td>
<td>Poor</td>
<td>00</td>
<td>00</td>
<td>108</td>
</tr>
</tbody>
</table>

Table 6-3: Comparison of Predictions: Cracking Related LV (MODEL2X and MODEL4)
The methodology for testing the contribution of a measurement to the modeling of deterioration therefore is somewhat more complicated than merely examining the squared multiple correlations of the structural and the measurement models. A broad outline of the procedure to be used can be described as follows:

1) Introduce the new measurements and calculate the squared multiple correlation of the structural equations and the measurement equation

2) If there is no major change in the fit of the deterioration equation (i.e., there is neither a large decrease or a large increase), and the squared multiple correlations of the new measurement equations are small, then it may be concluded that the new measurements do not significantly affect the latent variable, and may be dropped in the future. There should also be no major change in the parameter estimates of the structural model when the new measurements are included.

3) If there is an improvement in the fit of the structural equation, and the new measurements have a high squared multiple correlation, then the measurements contribute positively to the determination of the latent variable and should be included. It would also be a good test to check the correlation of the extracted values of the latent variable before and after the inclusion of the measurements. If the two measurements are contributing to the same latent variable as before, then the correlation between the extracted values should be high. It might even so happen that these new measurements explain the latent variable better than the existing measurements; at this stage, therefore, it might be useful to check if any of the other measurements can be dropped.

4) If there is a decrease in the fit of the structural equation, or the parameter estimates change
substantially, but the squared multiple correlation of the new measurements are high, then this may indicate that the new measurements are capturing the effect of another latent variable. The correlation between the extracted value of the new latent variable and the latent variable without the new measurements may not be very high. In this case, if the only latent variable of interest is the first one, then the measurements should be dropped; however, a better approach would be to include the new measurements (since they contain some information) and formulate a two latent variable model.

6.2 Prediction of Future Pavement Condition

Once the parameters of a deterioration prediction model are estimated, the model can be used to predict the condition of the pavement given a set of values of the explanatory variables. Fresh damage measurements may also be available, and this information can also be incorporated to produce improved predictions of future condition. Each scenario is now described.

If no new damage information is available, then all predictions are made from the structural model. Two cases are possible. The first scenario is where the future condition is to be predicted for a given extent of maintenance activities and explanatory variables. Such a scenario would occur when a highway agency is comparing different possible future maintenance activities and wants to know the effects of these activities. The predictions can be made simply by substituting the extents of maintenance and the explanatory variables into the deterioration model. The second scenario is where future condition is to be predicted for a given value of exogenous explanatory variables, but the maintenance is allowed to remain endogenous in response to changes in condition. The predictions are then made from the reduced form of the deterioration equation. The reduced form is obtained by substituting the relevant maintenance equation for the endogenous maintenance terms in the deterioration equation.
The predictions obtained for this scenario from the different models estimated in Chapter 5 are now compared. Three models are chosen for comparison; they are MODEL1, MODEL2 and MODEL4 from Chapter 5. MODEL1 is the simultaneous equation specification of the basic PSI model, where PSI is the performance index and the interactions between maintenance and deterioration are taken into account. MODEL2 is the "PSI like" latent variable model, where the same measurements that are used to construct the PSI are used in the measurement equations for the calculation of the latent variable. Finally, MODEL4 is the two latent variable model, where a deterioration equation is developed for both a "roughness related" and a "load or cracking related" latent variable. Only the deterioration equation for the roughness related latent variable is used for comparison here.

Comparisons are first made for the case where the maintenance variables are endogenous. The comparison is made on the basis of the elasticities of predicted condition from the three models with respect to changes in the explanatory variables. Table (6-4a) shows the percentage change in condition as the result of an increment of one year in age predicted by each model. The calculation is done for three different values of age; 5, 10 and 15 years. As is seen from the table, the PSI is more sensitive to changes in age than the other two models. Table (6-4b) shows the elasticity of predicted condition with traffic at three traffic levels; 500, 1000 and 2000 vehicles per day. The results indicate that the roughness related latent variable from MODEL4 is more sensitive to changes traffic than the latent variable from MODEL2. The PSI model gives surprising results. The elasticity of (-PSI) with respect to traffic is negative, indicating that the condition improves as traffic increases. The reason for this is that the reaction of the highway agency to decreasing condition is to over compensate by performing more maintenance. This can be seen from Fig (6-1), where condition (expressed as PSI in units of log(slope variance)) is plotted against traffic with and without maintenance being performed. Without maintenance, the condition of the pavement deteriorates with traffic; however, the
<table>
<thead>
<tr>
<th>AGE (years)</th>
<th>MODEL1</th>
<th>MODEL2</th>
<th>MODEL4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.47</td>
<td>2.32</td>
<td>1.61</td>
</tr>
<tr>
<td>10</td>
<td>1.64</td>
<td>1.11</td>
<td>0.37</td>
</tr>
<tr>
<td>15</td>
<td>0.25</td>
<td>0.22</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Table 6-4(a): Sensitivity of condition to 1-yr increase in age

<table>
<thead>
<tr>
<th>TRAFFIC (ADT)</th>
<th>MODEL1</th>
<th>MODEL2</th>
<th>MODEL4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.008</td>
<td>0.003</td>
<td>0.011</td>
</tr>
<tr>
<td>500</td>
<td>-0.016</td>
<td>0.007</td>
<td>0.022</td>
</tr>
<tr>
<td>2000</td>
<td>-0.054</td>
<td>0.011</td>
<td>0.065</td>
</tr>
</tbody>
</table>

Table 6-4(b): Sensitivity of condition to 1% increase in ADT
Figure 6-1: (-PSI) against traffic

- Condition (log(slope variance))
- Traffic (ADT)

- without maintenance
- with maintenance
maintenance performed to compensate for this deterioration causes the pavement to be in a better condition as traffic increases. Similar comparisons can be performed between the models on their sensitivity to other explanatory variables as well.

Figure 6-2 plots the predicted pavement performance as a function of age for each of the above three models. The pavement is chosen to be in district 3, with a road surface dummy taking the value 1 (i.e. flexible overlay on flexible pavement). All other explanatory variables are chosen to be at their mean values. Since the model is estimated using data in the form of deviations from the mean, this implies that all the explanatory variables have a value of zero. Predictions are made for age ranging from 0 to 25 years. All three performance measures are expressed in terms of the logarithm of the slope variance. As can be seen from the figure, the pattern of prediction of condition by the three models is very similar, except that the PSI model systematically predicts a worse condition than the other two models.

All models show an improvement in condition as the age of the pavement approaches 17 or 18 years. This arises from the fact that the effect of age is parabolic, and the function crosses the axis at a certain value of age. Predictions in this region are probably not realistic, though, since maintenance is endogenous, it is possible that the agencies are performing more maintenance as the pavement deteriorates, and this added maintenance over-compensates for the effect of age. What the results do indicate, however, is that the effect of age decreases as the pavement ages. This is also seen by the elasticities calculated in Table 6-4(a). The observation that the age effect decreases as the pavement ages is consistent with finding in the literature that the overall rate of pavement deterioration decreases as the pavement ages [as evidenced by the S-shaped performance function modeled by Riggins and Garcia-Diaz [1984], for example].
Figure 6.2: Deterioration Prediction

Comparison of Models

Condition (log(slope variance))

Age (years)

Model 1

Model 2

Model 3

Model 4
Of greater interest than the above comparisons is the predictions made by the different models when the maintenance is assumed to be exogenous. This scenario is more common for the purposes of maintenance planning, since it allows the comparison between different pre-specified maintenance activities. Fig (6-3) plots the predictions made by each model, where the extents of maintenance of sand seal coating, crack filling and chip seal coating are assumed to be at their mean levels. This corresponds to 807 sq. yds. of sand seal coating, 176 lbs. of crack filler and 569 sq. yds. of chip seal coating per pavement section. Fig (6-3) shows substantial differences between the predictions made by the three models. Under the "average maintenance" strategy, the PSI model underpredicts deterioration at low values of age, while overpredicting deterioration at high values of age compared to the latent variable models. The two latent variable models make similar predictions, but the roughness-related latent variable of the two variable model consistently predicts worse condition than the one variable model. Once again, the effect of age on condition decreases as age increases.

The difference between Figure (6-3) and Figure (6-2) can be explained in terms of the differences in the maintenance response of the agency to a unit change in condition for each performance measure. This response, calculated as the cumulative change in maintenance (i.e. summed over all three activities) for a unit increase in deterioration takes the values of 0.185, 0.036 and -0.010 for the PSI model, the single latent variable model and the two latent variables model respectively. This indicates that the agency performs more maintenance for a unit decrease in condition as measured by the PSI than for a unit decrease in condition measured by the latent variables. In fact, for the roughness related latent variable from the two variable model, the cumulative maintenance performed decreases for a unit increase in roughness. This arises because the extent of chip sealing decreases as the cracks get bigger. The result of these differences in maintenance response is that the performance measure with a more rapid deterioration rate (i.e. PSI) also has a larger maintenance response. This causes the three curves
Figure 6-3: DETERIORATION PREDICTION

Comparison of Models

Condition (log(slope variance))

Age (years)

- PSI
- Model2
- Model4
to move closer to one another in Figure (6-2).

The variance of the predicted performance variable, \( \text{Var}(S \mid A, X_1) \) for the predictions of Figure (6-3) can be calculated as follows:

The predicted value of \( S \), given the extent of maintenance and the explanatory variables, \( (\hat{S} \mid A, X_1) \) is given by:

\[
(\hat{S} \mid A, X_1) = \beta_1 A_1 + \beta_2 A_2 + \beta_3 A_3 + \beta_4 X_1
\]

(6-3)

Consider a single maintenance activity \( A_1 \) and two explanatory variables, Age and ADT. Then,

\[
(\hat{S} \mid A_1, \text{Age}, \text{ADT}) = \beta_1 A_1 + \beta_{41} \text{Age} + \beta_{42} \text{ADT}
\]

(6-4)

\[
\text{Var}(\hat{S} \mid A_1, \text{Age}, \text{ADT}) = A_1^2 \text{Var}\beta_1 + \text{Age}^2 \text{Var}\beta_{41} + \text{ADT}^2 \text{Var}\beta_{42} +
2A_1 \text{AgeCov}(\beta_1, \beta_{41}) + 2A_1 \text{ADTCov}(\beta_1, \beta_{42}) + 2\text{AgeADTCov}(\beta_{41}, \beta_{42})
\]

(6-5)

where Cov(a,b) stands for the covariance between a and b.

The variances and covariances of eqn (6-5) can be substituted for by their estimates obtained from the model. Tests for whether the differences between the predicted values by different models are significant can be performed from confidence intervals derived from these variances.

The above discussion has focussed on situations where information is available on the explanatory variables, but no fresh damage information exists. If damage measurements are
available, then these measurements should be used as well to provide improved predictions of condition. If damage measurements exist on new damage types that have not been included in the original estimation, then it is necessary to re-estimate the parameters using this damage information as well. The choice of whether this damage measurement should be included should be based on the criteria described in Section 6.1. If no new damage types are present, but data has been collected on the damage types originally used for estimation, then this information can be included for the prediction of condition by using the regression method factor extraction procedure described in Chapter 4. The extracted value of the latent variable $\hat{S}$ for a vector of explanatory variables $X$ and indicators I is given by the following general expression:

$$\hat{S} | I, X = BX + f(\lambda, \psi, \theta)(I - \lambda BX)$$  \hfill (6-6)

where $f(\lambda, \psi, \theta)$ is a function of the parameters. The exact expression for eqn (6-6) was presented in equation (4-22). Without the information on the indicators, the predicted value of $S$ is given by the first term of equation (6-6), i.e. $BX$. The second term uses the additional information that is available as well.

The issue of the differences in the predicted value of condition by different models is important because ultimately the predictive models are used for the determination of maintenance strategies. The choice of maintenance strategy depends upon the particular performance measure that is selected. For example, a simple maintenance rule that states that the pavement should be maintained when it is ten years old leads to different levels of roughness at which the pavement will be maintained, depending on the particular model chosen. From Figure 6-1, it is seen that the choice of either latent variable model predicts that the pavement being maintained at a lower roughness level than if the PSI model is chosen for the maintenance
decision. Conversely, if the maintenance decision is based upon a particular roughness level, use of the PSI model would cause the pavement to be maintained earlier than if any of the latent variable models were used. Clearly, this affects the allocation of resources for pavement resources over time and over the highway network. The choice of a suitable performance criterion and the corresponding performance prediction model is therefore intimately connected with the decision of a maintenance strategy.

The choice of a particular deterioration model or a performance criterion depends upon the specific problem, and it is difficult to set down general rules. Some criteria for selection are the characteristics of the performance measures (for example, the latent variable of MODEL2 is a stronger measure of ride quality than the PSI), the fit of the associated structural and measurement components of the different models and the variance of the predicted condition. Once a model has been chosen, the predictions of condition obtained from the model can be used to evaluate different maintenance options.
CHAPTER 7
DETERIORATION MODELS WITH DISCRETE ORDINAL INDICATORS

The models developed in Chapter 5 and 6 using the methods described in Chapter 4 estimated a deterioration model from measurements or assessments of damage on a continuous scale. In this chapter, a methodology for formulating deterioration models is described when the indicators are not measured on a continuous scale, but are instead measured on an arbitrary discrete ordinal scale (for example from 1 to 10) where 1 might be indicate a completely deteriorated facility and 10 might indicate a new facility. These discrete values indicating the condition of the facility are referred to as condition codes.

As the facility deteriorates, it is observed during successive inspections to pass through lower and lower condition codes. A deterioration model estimated from data of this sort has to be able to map in some manner the underlying latent continuous process of deterioration to the discrete transitions between condition codes observed during inspections. The terminology state, or condition state, is used interchangeably with condition code. In this chapter, several models of this sort are formulated. In this thesis, the models have not been implemented on real data because of the difficulty of obtaining such data; it is anticipated that this can be accomplished in future research.

A practical situation in which deterioration data is collected in such a manner is the case of concrete bridges. The condition of the major bridge components, such as the deck, superstructure or substructure is evaluated on a scale from 0 to 9, with 0 being the worst possible condition and 9 being the best [Seymour, 1985]. The three major mechanisms of deterioration for bridge decks are (a) scaling (b) cracking and (c) spalling. Scaling is the wear of low quality concrete as water freezes and thaws in small cracks and capillaries in the concrete. Cracking occurs primarily due to concrete

168
volume changes. Slow placement of concrete, excessive water-cement ratio, improper curing, high temperature during hardening and the subsidence of the concrete around the top reinforcement result in shrinkage which is the direct cause of cracking. Spalling is the corrosion of steel reinforcement resulting from the infiltration of corrosive elements such as salt used in deicing the deck. Salt destroys the thin protective iron-oxide layer on the surface of the steel bars. The electric potential of the reinforcing steel changes, anodic-cathodic cells form as a result of variations in chloride concentration along the bars and the anodic areas begin to rust. Table 7-1 shows the bridge deck condition code corresponding to different levels of these damage components.

The major difference between the problem considered in this chapter and the problem described in Chapters 3, 4 and 5 lies in the way deterioration is measured. Consider a simple latent variable causal model from Chapter 3, where a single latent variable can be expressed as a function of a vector of perfectly measured exogenous explanatory variables. The causal model can then be written as follows:

\[
S = \beta X + \eta
\]  

(7 - 1a)

where \( S \) is a latent performance variable, and \( X \) is a vector of explanatory variables. As discussed in Chapter 3, \( \beta \) is a vector of parameters and \( \eta \) is a vector of error terms. In the general linear latent variable formulation of Chapter 3, the corresponding measurement equation system for \( S \) can be written as follows:

\[
I = \lambda S + \varepsilon
\]  

(7 - 1b)
<table>
<thead>
<tr>
<th>Category Classification</th>
<th>Rating</th>
<th>Spalls</th>
<th>Delamination</th>
<th>Electrical Potential</th>
<th>Chloride Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category #3</td>
<td>9</td>
<td>none</td>
<td>none</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Light Deterioration</td>
<td>8</td>
<td>none</td>
<td>none</td>
<td>none &gt;0.35</td>
<td>none &gt;1.0</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>none</td>
<td>&lt; 2%</td>
<td>45% &lt;0.35</td>
<td>none &gt;2.0</td>
</tr>
<tr>
<td>Category #2</td>
<td>6</td>
<td>&lt; 2% spalls or sum of all deteriorated and/or contaminated deck concrete &lt; 20%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate Deterioration</td>
<td>5</td>
<td>&lt; 5% spalls or sum of all deteriorated and/or contaminated deck concrete 20 to 40%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category #1</td>
<td>4</td>
<td>&gt; 5% spalls or sum of all deteriorated and/or contaminated deck concrete 40 to 60%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extensive Deterioration</td>
<td>3</td>
<td>&gt; 5% spalls or sum of all deteriorated and/or contaminated deck concrete &gt; 60%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Deck structural capacity grossly inadequate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structurally Inadequate Deck</td>
<td>1</td>
<td>Deck has failed completely - Repairable by replacement only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>Holes in deck - danger of other sections of deck falling</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7-1: Rating Scheme for Bridge Deck Evaluation
where I is a vector of indicators. In the case of discrete indicators, the measurement equations (7-1b) must now be replaced by their discrete analog, since deterioration is measured as a series of transitions from one discrete condition state to a lower state. If damage is measured on a continuous scale, then one (observed) value of the indicator variable corresponds to one (unobserved) value of the latent variable. In the case of discrete states, however, one value of the indicator (i.e. a particular value of the condition code) corresponds to a range of values of the latent variable, and a transition from one condition state to the next takes place when the latent performance value passes outside the range. The range is specified by threshold values of the latent variable, which are unobserved, and can be estimated from the observed data of transitions over time. For example, the facility can be specified to remain in condition code 9 as long as S lies between the values $S^*_9$ and $S^*_8$. As long as the latent performance vector remains between any two specific threshold values, the facility is observed to remain in the same state. Fig 7-1 shows the process of deterioration both in the latent variable space and in the indicator space over time.

In equation (7-1), S is considered to be a one-dimensional scalar. In general, this does not need to be the case. Consider, for example, the condition of the bridge as a whole. The bridge consists of many components; deck, superstructure, substructure etc. The condition of each of these components is latent. S is therefore a 3-dimensional vector, and correspondingly there exists a 3-dimensional state vector of indicators. The state of the bridge is then characterized by the value of the indicator along each dimension. For example, the bridge will be in state $(9,9,9)$ as long as the associated latent variables $S_1, S_2, S_3$ each lie between certain pre-specified thresholds. It is also possible to have a single indicator for multiple latent variables. In such a case, the facility is observed to be in some state i if, for example, $S_1$ lies between $S^{*}_{19}$ and $S^{*}_{18}$, and $S_2$ lies between $S^{*}_{29}$ and $S^{*}_{28}$.
Figure 7-1: Correspondence between Condition and States
and \( S_3 \) lies between \( S_{39}^* \) and \( S_{38}^* \). Such a representation is not very realistic because it compresses information about a variety of processes into a single number. In the case of bridges, a separate condition code is given to each bridge component.

A facility that deteriorates from some state \( i \) to a lower state \( j \) in \( t \) time periods successively passes from state \( i \) to the next lower state \( i-1 \) in a certain number of time periods, to the next lower state \( i-2 \) in some more time periods, and so on until state \( j \) is reached. The observed process of deterioration therefore corresponds to a sequence of "holdings" in a given state and "transitions" from this state to the next lower state whenever the underlying process crosses a threshold.

Estimation of the deterioration model involves estimating the matrix of parameters of equation (7-1a) and the set of thresholds for each latent variable from the sequence of observed transitions and data on a vector of explanatory variables affecting deterioration. For most infrastructure facilities, observation on condition is done through visual inspection conducted once every few years. In the case of bridge decks, for example, an inspection is performed every two years. As a result, complete information on the sequence of holdings and transitions is not available. All that is known is that at the point in time corresponding to the inspection, the facility was in a particular state; it is not known when the sequence of transitions occurred that brought the facility to this state. The only exception to this is in the case when the facility is new or has just been rehabilitated, because it is then known that the facility entered the "new" state at that time. In all other cases, inspection data simply provides information on the following event: the facility is in state \( i \) at one inspection and is in state \( j \) at the next inspection \( t \) time periods later. A typical cross-sectional data set contains observations of such events for many facilities. In order to write the likelihood function to estimate the parameters of the deterioration model using Maximum Likelihood techniques, it is
necessary to compute the probability of the observed transition for each facility in the data set. The
product of these probabilities is the likelihood function. In this and the following sections, the
methodology for calculating this probability is developed.

Because of the complexity of some of the specifications, the deterioration model considered in
this chapter is the simplest version of the general latent variable model formulated in Chapter 3.
All explanatory variables are assumed to be perfectly measured and exogenous. Also, the simulta-
neity effects between deterioration and maintenance have not been included in these models. The
assumption under which the simultaneity does not exist is that the maintenance policy is exogenously
specified and is not triggered by the condition of the facility. This is a more reasonable assumption
for bridge decks than for highway pavements because the number of damage types is smaller for
bridge decks than for highway pavements and the number of possible responses to damage is
correspondingly small as well. Therefore, the maintenance activities over time are relatively
constant and can be better approximated by an exogenous policy than in the case of bridge decks.

Another assumption of the model is that the threshold values described above are deterministic.
This is equivalent to assuming that there is no error in the measurement of the condition state; i.e.
the measurement equations are error free. Finally, since the condition indices are aggregate measures
of various damage types (see Figure 7-1), only one indicator exists for each latent variable. This
is equivalent to a situation in the models of the previous chapters where instead of the different
damage measurements, the PSI or the PCI alone are used as indicators for the latent variable.

The methodology for the calculation of the probability of the observed transition is developed in
this chapter in three stages. The first stage concentrates on a single time period and calculates
expressions for the probability of a transition to a lower state in this time period. The complement of this probability is the probability that no transition occurs in the time period, i.e. facility "holds" in the same state. The second stage uses the single time period probabilities to calculate the probability of one transition. Since the facility may hold in a state for several time periods before undergoing a transition this probability is the joint probability of holding for multiple periods and then undergoing a transition to a lower state. Clearly, this probability can be expressed in terms of the single time period probabilities of the first stage. Finally, the observed deterioration event in the data for each facility is a transition from some state i to a lower state j and can consist of several transitions. In the third stage, the probability of these "multi-step" transitions are built up from the single step transitions of the second stage. A flow-chart for this process is illustrated in Fig. 7-2.

7.1 State transitions in a single time period

In this section, expressions of the first stage of the process illustrated in Fig 7-2 is derived. Define $S(t)$ as a (P x 1) vector of latent performance variables at time $t$. The elements of $S(t)$ are denoted by $S_j(t)$, $j=1,2,...,P$. The P performance variables relate to the performance characteristics of the facility, e.g. ride quality, structural strength etc. Equation (7-1a), repeated below, explains the progression of the latent performance characteristics as a function of the factors affecting deterioration:

$$S(t) = \beta X(t) + \eta(t)$$  \hspace{1cm} (7 - 1a)

where

$\beta$ is a (P x K) matrix, $X(t)$ is a (K x 1) vector, $\eta(t)$ is a (P x 1) vector.
probability of moving to next lower state in current time period

probability of holding in state i for n time periods and then moving to next lower state

probability of moving from state i to lower state j in m time periods

Figure 7-2 : Calculation of Transition Probabilities
Let $I(t)$ be the $(P \times 1)$ vector of observed indicators at time $t$. Assume that a single indicator exists for each latent variable, so the dimension of the vector $I$ and the vector $S$ are the same. The elements of $I(t)$ can be written as $I_1(t), I_2(t), \ldots, I_P(t)$ and each element is an indicator of one element of $S(t)$. Each element of $I(t)$ can take on a number of discrete values reflecting the observed condition of the facility. The condition of the facility at any point in time is therefore specified by the $P$ discrete values taken by the elements of vector $I$.

The connection between the indicator and the latent variable, which is the discrete analog of equation (7-1b), can be written as follows for each latent variable and its associated indicator as follows:

$$I_j(t) = i \Rightarrow S_j^{i-1} \leq S_j(t) < S_j^i \quad j = 1, 2, \ldots, P$$

where:

$I_j(t)$ is the $j$th indicator at time $t$

$i$ is the value (i.e. the condition state) of the $j$th indicator at time $t$

$S_j(t)$ is the $j$th latent variable at time $t$

$S_j^{i-1}, S_j^i$ are the threshold values of the $j$th latent variable.

Equation (7-2) states that the $j$th indicator of facility condition is observed to be in state $i$ at time $t$ if the $j$th latent variable lies between threshold values $S_j^{i-1}$ and $S_j^i$. In equation (7-2), it is assumed that state $i-1$ represents a worse condition than state $i$ and that $S_j^i > S_j^{i-1}$ (i.e. smaller values of the latent variable implies a poorer condition). When $X(t)$ is zero, then the value of $S(t)$ is zero, it's maximum value; deterioration produces successively more and more negative values of $S(t)$.
The first stage in the process of Figure 7-2 is the calculation of the transition probability for a single time period. In any time period, except in a period immediately after reconstruction or rehabilitation, this probability in general depends upon the sequence of transitions over time. There are several reasons for this dependence. The first reason is the temporal correlation between the error terms of the latent performance equation (for example between \( \eta_j(t) \) and \( \eta_j(t-1) \) for any \( j \)). This kind of correlation can occur for a variety of reasons. If deterioration is affected by some unobserved exogenous characteristic (e.g. some weather related phenomenon) that is captured by the error term, this characteristic might persist across adjacent time periods. Another cause for serial correlation may be the presence of time independent unobserved characteristics that persist over time through the lifetime of the facility. An example of such a characteristic could be the initial quality of construction of the facility. In this context, the information that \( \eta_j \) had a particular value at time \( t-1 \) conveys information about the value of \( \eta_j \) at time \( t \). To illustrate this further, consider the following modification equation to (7-1a) for performance variable \( j \) in time period \( t-1 \), where the exogenous variable \( X \) has been removed:

\[
S_j(t-1) = \beta_j + \eta_j(t-1)
\]  

(7 - 3)

In equation (7-3), the condition at time \( t-1 \) therefore depends only upon the error term \( \eta_j(t-1) \) and a time independent constant \( \beta_j \).

Suppose the \( j \)th indicator of the facility \( I_j \) is observed to have the value \( i \) in time period \( t-1 \). Then substituting equation (7-3) in equation (7-2), we have:
\[ I_j(t-1) = i \Rightarrow S_j^i - 1 \leq \eta_j(t-1) < S_j^i \leq \beta_j \quad (7-4a) \]

Suppose the facility does not undergo a transition in time period t-1 so that in the next time period, it is still in state i. The analogous to equation (7-4), we have,

\[ I_j(t) = i \Rightarrow S_j^{i-1} \leq \eta_j(t) < S_j^i \leq \beta_j \quad (7-4b) \]

The conditional probability for the above event, i.e. that the jth indicator is observed in state i at time t given that it was in state i-1 at time t-1 can then be written as follows:

\[
\begin{align*}
\text{Prob}(I_j(t) = i \mid I_j(t-1) = i) = \\
\text{Prob}(S_j^{i-1} - \beta_j \leq \eta_j(t) < S_j^i - \beta_j \mid S_j^{i-1} - \beta_j \leq \eta_j(t-1) < S_j^i - \beta_j) \\
\quad (7-5)
\end{align*}
\]

From Bayes’ law, the conditional probability of eqn (7-5) depends upon the joint probability distribution of \( \eta_j(t) \) and \( \eta_j(t-1) \). If the error terms are independent, then the above conditional probability does not depend on \( \eta_j(t-1) \) and is simply given by:

\[
F(S_j^i - \beta_j) - F(S_j^{i-1} - \beta_j) \quad (7-6)
\]

where \( F(\cdot) \) is the cumulative distribution function of \( \eta_j(t) \)

In general, the conditional probability of equation (7-5) increases monotonically with the correlation between \( \eta_j(t) \) and \( \eta_j(t-1) \). If the error terms are uncorrelated, then the probability of a transition.
in time period $t$ is independent of a transition in time period $t-1$ as in equation (7-6). If the two terms are perfectly correlated, then if $I_j(t - 1) = i$, then it is certain that $I_j(t) = i$.

Dependence on the past can exist even without temporal correlation of the error terms. If the condition at time $t$ in the deterioration model of equation (7-1a) is expressed as a function of cumulative explanatory variables such as cumulative traffic or age since the construction of the facility, then the probability of transition to a lower state increases with the duration of time the facility has spent in the current state. This is seen by rewriting equation (7-4) including exogenous variables. The error terms are assumed to be independent to remove any temporal correlation effects. For time period $t-1$,

$$
Prob(I_j(t - 1) = i) = Prob(S_j^{i-1} - \beta_j X_j(t - 1) \leq \eta_j(t - 1) < S_j^{i} - \beta_j X_j(t - 1)) \quad (7-7a)
$$

Similarly, for time period $t$,

$$
Prob(I_j(t) = i) = Prob(S_j^{i-1} - \beta_j X_j(t) \leq \eta_j(t) < S_j^{i} - \beta_j X_j(t)) \quad (7-7b)
$$

If $X_j$ is cumulative, $\beta_j X_j(t)$ is greater in absolute value than $\beta_j X_j(t - 1)$. Since both $S_j^{i-1}$ and $S_j^{i}$ take on negative values, in equation (7-7a) a smaller amount is added to the thresholds than in equation (7-7b). If we assume that $\eta_j(t)$ and $\eta_j(t - 1)$ have the same distribution, then the probability of remaining in the same state at time $t$, given by equation (7-7b) is smaller than the probability of remaining in the same state at time $t-1$, given by equation (7-7a). The length of time spent in the state therefore influences the probability of future transitions.
Summarizing, the following points can be made about the nature of a transition from a given state to a lower state in a single time period:

(1) The observed discrete transitions from state to state take place when the underlying latent performance variable crosses some thresholds. These thresholds are unobserved and need to be estimated from the data along with the parameters of the deterioration model.

(2) The stochastic nature of the deterioration of the facility can be characterized in the indicator space by the probability of transition from one state to a lower state during a given time period.

(3) The probability of transition in a given time period depends upon the history of past transitions if there is a temporal correlation between the error terms of the deterioration equation.

(4) Even if no temporal correlation is present, dependence on history can still be present if the deterioration model is specified using cumulative values of the explanatory variables.

The above points lay the framework for developing models for calculating single period transition probabilities. The methodology is presented for the case of a one dimensional latent variable $S(t)$ at time $t$. The indicator for this variable is also one dimensional and is denoted $I(t)$ at time $t$. In this case, the system of $P$ equations represented in equation (7-1a) reduces to a single equation and the state is represented by a single value of $I(t)$. The one dimensional analog of equation (7-2) establishes the relation between the latent variable and its indicator at time $t$. Suppose the indicator $I(t)$ is in state $i^*$. We then have:
\[ I(t) = i^* \Rightarrow S^{i^* - 1} \leq S(t) < S^{i^*} \quad (7-2a) \]

where \( S^{i^* - 1} \) and \( S^{i^*} \) are the threshold values. Suppose the facility is new, or just rehabilitated at time \( t_0 \). It is assumed at this time that the facility is in the highest possible state. Over time, it is assumed that the facility drops to lower and lower states; a transition to a higher state is not possible unless rehabilitation is performed. This is equivalent to assuming that routine maintenance performed between rehabilitations serves only to retard the rate of deterioration but does not cause any positive improvement. This has been pointed out by many authors in the infrastructure literature [see, for example, Markow, 1984].

Consider next a time \( t+1 \) which is one time interval after time \( t \). The choice of the length of the time interval is crucial to the derivations which follow. Assume that the interval of time between \( t \) and \( t+1 \) is chosen small enough such that if a transition does take place in the interval, only one transition can take place so that the only state that can be reached is the one that is immediately below the current state. In other words, if the value of \( I(t) \) is \( i^* \), \( I(t+1) \) can only take the values of \( i^* \) or \( i^* - 1 \) where \( i^* - 1 \) is the next lower state. This "Poisson process" like assumption is central to the model. For convenience, it is also assumed that if a transition does take place in a particular interval, it occurs at the beginning of the interval.

Consider a time \( t_0 + n - 1 \) which is \( n-1 \) time periods after rehabilitation or construction of the facility. Let \( I(t_0 + n - 1) = i^* \). Suppose a transition takes place at the beginning of the next interval. Then, by definition of the time interval, \( I(t_0 + n) = i^* - 1 \). The probability of this transition is the "single time period" transition probability which will be used as a building block in the derivations that follow. For reasons already discussed, this probability depends both upon the present state as well as on
the sequence of past transitions that took place before this state was reached. It is therefore defined as the **conditional probability of transition**, where the conditioning is on the sequence of events till $t_0+n-1$. This sequence of transitions till time $t_0+n-1$ is denoted by $H(t_0+n-1)$ and is the **history** of the deterioration process before the present time. Assuming $k$ transitions from the new state to the present state $i^*$,

$$H(t_0+n-1) = \{t_0+n_1, t_0+n_2, \ldots t_0+n_k \mid (t_0+n-1) = i^*\}$$

where $n_1, n_2, \ldots, n_k$ are the intervals after $t_0$ at which transitions take place.

The **conditional probability** of a transition in time period $t_0+n$ **given** the history of past transitions can then be defined as follows:

$P(i^*, i^{*}-1 \mid I(t_0+n-1), H(t_0+n-1))$ is the conditional probability of transition from state $i^*$ to state $i^{*-1}$ in time period $t_0+n$, given the state at time $t_0+n-1$ and the sequence of transitions from the last rehabilitation or construction to time period $t_0+n-1$.

From eqn (7-2a), this probability can be expressed as:

$$P(i^*, i^{*}-1 \mid I(t_0+n-1), H(t_0+n-1)) =$$

$$\text{Prob}(S^{i*-2} \leq S(t_0+n) < S^{i*-1} \mid S^{i*-1} \leq S(t_0+n-1) < S^{i*}, H(t_0+n-1)) \quad (7-8)$$

The complementary event is that no transition takes place in time period $t_0+n$ and the facility remains in state $i^*$ in this period. The probability of this event is denoted by $P(i^*, i^* \mid I(t_0+n-1), H(t_0+n-1))$. This probability can be written as:
\[ P(i^*, i^* - 1 | I(t0 + n - 1), H(t0 + n - 1)) = 1 - P(i^*, i^* - 1 | I(t0 + n - 1), H(t0 + n - 1)) \]  \( 7 - 9 \)

Equations (7-8) and (7-9) respectively define the probabilities of transition and holding respectively for a single time period \( t0+n \). More specifically, equation (7-8) can be written in terms of the parameters of the deterioration equation by substituting equation (7-1a) into equation (7-8) as follows:

\[ P(i^*, i^* - 1 | I(t0 + n - 1), H(t0 + n - 1)) = \]

\[ Prob(S^{i^* - 2} \leq \beta X(t0 + n) + \eta(t0 + n) < S^{i^* - 1} \leq \beta X(t0 + n - 1) + \eta(t0 + n - 1) < S^{i^*}, H(t0 + n - 1)) \]  \( 7 - 10 \)

Different deterioration models can be developed from equation (7-10) by making different assumptions about the dependence of future transitions on the past, i.e., assumptions about the history term \( H(t0+n-1) \). Several models are now derived in the following section.

7.2 Some models of facility deterioration

A. Simple Bernoulli (independent error term) model:

This is the simplest possible stochastic model. In this model, the error term of the deterioration model is uncorrelated across time periods. In addition, no exogenous variable vector is included in the deterioration model. There is therefore nothing to distinguish one observation from another; the deterioration of each facility is completely random. The deterioration model for the latent
condition at time $t$ is given by the following equation:

$$S(t) = \beta_0 + \eta(t) \quad (7-11)$$

In such a model, the condition at time period $t$ depends only on the value of the error term; if $\eta(t)$ is large enough, then a transition takes place; otherwise the facility remains in the same state. The process of deterioration over time is therefore a sequence of Bernoulli trials. The conditional probability of transition from state $i^*$ to $i^*-1$ in any time period is independent of history.

Therefore, in this case eqn (7-10) can be written as follows, with the conditioning on history removed:

$$P(i^*, i^*-1 \mid I(t0 + n - 1), H(t0 + n - 1)) =$$

$$\text{Prob}(S^{i^*-2} \leq S(t0 + n - 1) < S^{i^*-1}) =$$

$$\text{Prob}(S^{i^*-2} - \beta_0 \leq \eta(t0 + n - 1) < S^{i^*-1} - \beta_0) \quad (7-12)$$

If $\eta$ is identically distributed across time periods, then this probability is constant for each time period. Suppose $\eta$ is normally distributed with mean 0 and standard deviation $\sigma_\eta$, then eqn (7-14) becomes:

$$P(i^*, i^*-1 \mid I(t0 + n - 1), H(t0 + n - 1)) = \Phi \left( \frac{S^{i^*-1} - \beta_0}{\sigma_\eta} \right) - \Phi \left( \frac{S^{i^*-2} - \beta_0}{\sigma_\eta} \right) \quad (7-13)$$

where $\Phi$ is the cumulative standard normal distribution.
B. Simple Bernoulli model with heterogeneity

The above model assumes no variation in characteristics across time or across facilities. In a slightly more complicated model, variation is introduced through a vector of exogenous variables \( X \). However, the error terms are still assumed to be independent across time periods. The deterioration equation in this case is given by equation (7-1a):

\[
S(t) = \beta X(t) + \eta(t) \quad (7-1a)
\]

where \( \eta(t) \) is independent across time periods. In this model, the conditional probability of transition is still independent of history, but the probability is different for different time periods and different facilities depending upon the value of the explanatory variable vector \( X \). An analogous equation to (7-13) can be written for this model that includes the vector of exogenous variables \( X \) as follows:

\[
P(i^*, i^* - 1 \mid I(t0 + n - 1), H(t0 + n - 1)) = \\
\Phi \left[ \frac{S_{i^*}^* - \beta X(t0 + n)}{\sigma_{\eta}} \right] - \Phi \left[ \frac{S_{i^*}^{*-2} - \beta X(t0 + n)}{\sigma_{\eta}} \right] \quad (7-14)
\]

Unlike equation (7-13), the probability of transition is not constant for each time period. As already discussed in the previous section, if \( X \) is a vector of cumulative explanatory variables, the probability of transition increases with the number of time periods already spent in the state.

C. A random effects model:
In this model, heterogeneity between observations is provided not only by the vector of exogenous variables, but also by a time persistent effect in the error term. The error $\eta(t)$ in any time period is the sum of two components: (i) a disturbance $\epsilon(t)$ with mean 0 and standard deviation $\sigma$ for each time period and (ii) a facility specific error term $\nu$ distributed with mean 0 and standard deviation $\sigma_\nu$ that is constant across time periods. This model therefore postulates an unknown and unmeasurable difference between facilities that affects the deterioration of the facility in addition to the intangibles that are present in each time period. Such time independent, facility specific unobservable characteristics could relate to the quality of the material with which the facility is built or the efficiency of the labor used in the construction of the facility.

Suppose it is assumed that $\epsilon(t)$ is uncorrelated across time periods. However, the presence of the time independent error term $\nu$ causes temporal correlation. The covariance across time periods is:

$$ Cov(\eta(t), \eta(t - 1)) = Cov(\epsilon(t) + \nu, \epsilon(t - 1) + \nu) = \sigma_\nu^2 \quad (7 - 15) $$

In the Bernoulli models discussed in the previous sub-section, this covariance is zero.

If it is assumed that the population of facility specific errors is distributed with pdf $f_\nu(\nu)$, then each facility has a fixed value of $\nu$ drawn from this distribution. Suppose that for one particular facility, the facility specific error is equal to some value $q$. Suppose, also, that the time dependent error component $\epsilon$ is normally distributed with mean 0 and standard deviation $\sigma_\epsilon$. The composite error term $\eta$ at time period $(t0+n)$ is equal to $\epsilon(t) + q$, where $q$ is a constant, i.e. $\eta(t0+n)$ is normally
distributed with mean $q$ and variance $\sigma_e$. **Given the value of $q$,** there is no difference between this model and the heterogeneous Bernoulli model of equation (7-14), and the single time period transition probability **given** $q$ can be written as follows analogous to equation (7-14):

$$P(i^*, i^* - 1 \mid I(t0+n-1), H(t0+n-1), q) =$$

$$\Phi\left[\frac{S^{i^*-1} - \beta X(t0+n) - q}{\sigma_e}\right] - \Phi\left[\frac{S^{i^*-2} - \beta X(t0+n) - q}{\sigma_e}\right]$$

(7 - 16)

The conditioning on $q$ in equation (7-16) can be removed to provide the average conditional transition probability at time $t0+n$ by integrating over the distribution of $v$. Suppose $v$ is also normally distributed with mean 0 and standard deviation $\sigma_v$. Then:

$$P(i^*, i^* - 1 \mid I(t0+n-1), H(t0+n-1)) =$$

$$\int \left\{\Phi\left[\frac{S^{i^*-1} - \beta X(t0+n) - q}{\sigma_e}\right] - \Phi\left[\frac{S^{i^*-2} - \beta X(t0+n) - q}{\sigma_e}\right]\right\}\phi\left\{\frac{q}{\sigma_v}\right\} dq$$

(7 - 17)

where $\Phi(\cdot \mid \cdot)$ and $\phi(\cdot \mid \cdot)$ are the CDF and the PDF of the standard normal distribution respectively.

Equation (7-17) gives the average probability that a facility with characteristics described by the exogenous variable vector $X(t0+n)$ undergoes a transition from $i^*$ to $i^* - 1$ in time period $t0+n$. The dependence across time periods arises from the presence of the term dependent upon the $v$ term, which is a time independent effect that will be present in the expression for the transition probability for each time period.
**D. Latent Markov Model**

All the models described so far have assumed that there is no direct dependence of the error term of one time period on the error term of another time period. Any correlation between time periods has arisen due to the presence of underlying time independent errors. The model described in this sub-section assumes a first order serial correlation between the error terms of successive time periods. \( \eta(t) \) and \( \eta(t - 1) \) are connected by the following relationship:

\[
\eta(t) = \eta(t - 1) + \varepsilon(t) \quad (7 - 18)
\]

where \( \varepsilon(t) \) is uncorrelated with \( \eta(t - 1) \)

Successively substituting for \( \eta(t) \) in equation (7-18), we get

\[
\eta(t) = \eta(0) + \sum_{k=0}^{t-1} \varepsilon(t - k) \quad (7 - 19)
\]

If \( \eta(0) \) is interpreted as the facility specific error \( \nu \) of the random effects model described above, then this model can be interpreted as an extension of the random effects model where the error at time \( t \) depends upon the facility specific error and the independent error terms of all previous periods.

Consider two time periods \( t0+n \) and \( t0+n-1 \) as before. From eqn (7-1a),

\[
S(t0+n) = \beta X(t0+n) + \eta(t0+n)
\]
\[ S(t0 + n - 1) = \beta X(t0 + n - 1) + \eta(t0 + n - 1) \]  

\[ (7-20) \]

Subtracting,

\[ S(t0 + n) - S(t0 + n - 1) = \beta[X(t0 + n) - X(t0 + n - 1)] + [\eta(t0 + n) - \eta(t0 + n - 1)] \]

If \( S(t0+n-1) \) is known and constant, then the above equation can be rewritten as follows:

\[ S(t0 + n) = S(t0 + n - 1) + \beta[X(t0 + n) - X(t0 + n - 1)] + \varepsilon(t0 + n) \]  

\[ (7-21) \]

where \( \varepsilon(t0 + n) \) is uncorrelated with the error term of other time periods by definition.

From equation (7-21), if the condition at \( t0+n-1 \) is known, then the condition at \( t0+n \) depends only upon the change in exogenous variables at time \( t0+n \) and an independent error term \( \varepsilon(t0 + n) \). This is a characterization of a Markov process [Ross, 1983].

This model is called a "latent Markov model" after Coleman[1964], because the Markov property applies only to the underlying latent deterioration equation. It does not apply to transitions in the indicator space. This is because the indicator space is an aggregation of a range of values from the underlying continuous latent variable space. The probability of transition in the indicator space depends upon how close the present value of the latent variable is to the corresponding transition threshold, i.e. it depends on how much time has already been spend in the present state.

To calculate the conditional probability of transition in a time period \( t0+n \), once again equation
(7-10) is used as the starting point, and the equation is repeated below:

\[ P(i^*, i^* - 1 | I(t0+n-1), H(t0+n-1)) = \]

\[ \text{Prob} \{ S^{i^*-2} \leq S(t0+n) < S^{i^*-1} | S^{i^*-1} \leq S(t0+n-1) \leq S^{i^*}, H(t0+n-1) \} \quad (7-22) \]

The equation can be rewritten as follows:

\[ P(i^*, i^* - 1 | I(t0+n-1), H(t0+n-1)) = \]

\[ \text{Prob} \{ S^{i^*-2} \leq S(t0+n-1) + \Delta S(t0+n) < S^{i^*-1} | S^{i^*-1} \leq S(t0+n-1) < S^{i^*}, H(t0+n-1) \} \quad (7-23) \]

where \( \Delta S(t0+n) \) is the change in condition in time period \( t0+n \).

If the value of \( S(t0+n-1) \) is known, then, supposing that this value is \( q \), equation (7-23) can be written as follows:

\[ P(i^*, i^* - 1 | S(t0+n-1) = q, H(t0+n-1)) = \]

\[ \text{Prob} \{ S^{i^*-2} - q \leq \Delta S(t0+n) < S^{i^*-1} - q | S(t0+n-1) = q \} \quad (7-24) \]

In equation (7-24), the probability of transition is now expressed in terms of the change in condition at time \( t0+n \). Since the underlying process is Markovian, the probability that \( \Delta S(t0+n) \) lies between some thresholds given \( S(t0+n-1) \) is independent of history. The history term has therefore been removed from the right hand side of equation (7-24).
Equation (7-24) is conditional on a specific value of the condition at time t0+n-1. If it is assumed that this value came from a distribution \( f_{S(t0+n-1)} \) of condition in the population, then the average probability of transition from state \( i^* \) to \( i^*-1 \) at time t0+n can be obtained by integrating over the pdf of \( S(t0+n-1) \) as was done in the random effects model of the previous sub-section. Substituting \( \beta \Delta X(t0 + n - 1) \) for \( \Delta S(t0 + n - 1) \), where \( \Delta X(t0 + n - 1) \) is the incremental change in the explanatory values in time period t0+n, the following is obtained:

\[
P(i^*, i^*-1 \mid S(t0 + n - 1) = q, H(t0 + n - 1)) = \\
Prob\{S^{i^* - 2} - q - \beta \Delta X(t0 + n) \leq \epsilon(t0 + n) < S^{i^* - 1} - q - \beta \Delta X(t0 + n)) \mid S(t0 + n - 1) = q\} \\
(7 - 25)
\]

If \( \ldots \) pdf of \( S(t0+n-1) \) is \( f_{S(t0+n-1)}(S(t0 + n - 1)) \), and the distribution function of \( \epsilon \) is \( F_\epsilon \), then the conditioning on \( S(t0+n-1) \) can be removed by integrating over all possible values of \( S(t0+n-1) \). Since \( S(t0+n-1) \) can only take values between \( S^{i^* - 1} \) and \( S^i \), these are the limits of integration. The average probability of transition from state \( i^* \) to state \( i^*-1 \) at time t0+n-1 is then

\[
P(i^*, i^*-1 \mid S(t0 + n - 1) = q, H(t0 + n - 1)) = \\
\frac{\int_{S_i}^{S_i} \left\{ [F_\epsilon(S^{i^* - 1} - q - \beta \Delta X(t0 + n))] - [F_\epsilon(S^{i^* - 2} - q - \beta \Delta X(t0 + n))] \right\} f_{S(t0+n-1)}(q) dq}{\int_{S_{i-1}}^{S_i} f_{S(t0+n-1)}(q) dq} \\
(7 - 26)
\]

where the denominator is an adjustment for the pdf to integrate to 1 since the integration is done
only over a subset of possible values of condition.

Equation (7-26) gives the average probability of transition from state $i^*$ to $i^*-1$ at time $t0+n$ for a latent Markov model. The pdf of $S(t0+n-1)$ can be derived from equation (7-19). If $\eta(0)$ is normally distributed with mean 0 and standard deviation $\sigma_\eta$ and the $\epsilon$ terms are normally distributed with means zero and standard deviation $\sigma_\epsilon$, then $S(t0+n-1)$ is normally distributed with mean $\int \lambda(t0 + n - 1)$ and a standard deviation that is easily calculated by summing the standard deviations of the error terms.

In this section, expressions for the one period transition probabilities in terms of the parameters to be estimated have been calculated for some special cases of the general model corresponding to specific assumptions about the structure of the error terms. Other models can be derived in a similar manner from other assumptions, though as can be seen from the models derived above, the algebra gets rapidly complicated. In the following section, we turn to Stages 2 and 3 of Fig. 7.2, and demonstrate how the single period transition probabilities can be used to calculate the joint probabilities of holdings and a single transition over multiple time period which in turn are used to derive expressions for the probabilities of transitions observed in the data. Maximum likelihood can then be used to estimate the parameters of the deterioration equation and the thresholds.

7.3 Multiple period transition probabilities:
Suppose the facility that reached state $i^*-1$ in the beginning of time period $t_0+n$ remains in this state till some future time $t_0+n^*-1$. In time period $t_0+n^*$ the facility undergoes a transition to state $i^*-2$. The facility therefore "holds" in state $i^*-1$ from $t_0+n+1$ to $t_0+n^*-1$, and then goes to a lower state $i^*-2$ in time period $t_0+n^*$. Define

$h_{i^*-1,i^*-2}(t_0+n+1, t_0+n^* \mid I(t_0+n+1), H(t_0+n))$ as the joint probability that the facility is in state $i^*-1$ from time $t_0+n+1$ to $t_0+n^*-1$ and then undergoes a transition to state $i^*-2$ at $t_0+n^*$, given the sequence of past transitions before $t_0+n$.

The event of a holding from $t_0+n+1$ to $t_0+n^*-1$ and then a transition at $t_0+n$, given the history of past transitions can be built up from: the intersection of the following events:

- no transition at $t_0+n+1$,
- no transition at $t_0+n+2$,
- no transition at $t_0+n+3$,
- . . .
- . . .
- . . .
- no transition at $t_0+n^*-1$,
- transition from $i^*-1$ to $i^*-2$ at $t_0+n^*$

The probability of this event defined above can be written as follows:

$$h_{i^*-1,i^*-2}(t_0+n-1, t_0+n^* \mid H(t_0+n+1), I(t_0+n+1)) =
\text{Prob}\{(\text{no transition at } t_0+n+1) \cap (\text{no transition at } t_0+n+2) \cap . . . \cap (\text{no transition at } t_0+n^*-1) \cap}$$
(transition from \(i^* - 1 \text{ to } i^* - 2 \text{ at } t0 + n^* \) \( | H(t0 + n), I(t0 + n + 1) \))

Employing Bayes' law, this joint probability can be written as the product of the following conditional probabilities:

\[
\begin{align*}
\text{Prob}(\text{no transition at } t0+n+1 | \text{transition at } t0+n, H(t0+n)), \\
\text{Prob}(\text{no transition at } t0+n+2 | \text{no transition at } t0+n+1, H(t0+n+1)), \\
\text{Prob}(\text{no transition at } t0+n+3 | \text{no transition at } t0+n+2, H(t0+n+2)), \\
\ldots \\
\ldots \\
\ldots \\
\text{Prob}(\text{no transition at } t0+n^*-1 | \text{no transition at } t0+n^*-2, H(t0+n^*-2)), \\
\text{Prob}(\text{transition at } t0+n^*-1 | \text{no transition at } t0+n^*-1, H(t0+n^*-2))
\end{align*}
\]

Examination of these probabilities reveals that they are nothing more than the one-step conditional probabilities derived in equations (7-8) and (7-9). The joint probabilities can therefore be constructed as the product of the these one-step probabilities as follows:

\[
h_{i^* - 1, i^* - 2}(t0 + n + 1, t0 + n^* | H(t0 + n + 1)) = \\
P(i^* - 1, i^* - 1 | I(t0+n), H(t0+n))P(i^* - 1, i^* - 1 | I(t0+n+1), H(t0+n+1)). \ldots \\
P(i^* - 1, i^* - 1 | H(t0+n^*-2), I(t0+n^*-2))P(i^* - 1, i^* - 2 | I(t0+n^*-1), H(t0+n^*-1))
\]

(7-27)

where the P(.|.) terms are the one-step conditional probability terms from equations (7-8) and (7-9).

The next step (stage 3 of Figure 7-2) is to use these expressions for the joint probability to build up expressions for the transitions observed in the inspection data. The expression for the joint
probability given by equation (7-27) is for one specific sequence of holdings and transitions, i.e.,
the event where the facility remains in state i*-1 from time period t0+n+1 to t0+n*-1 and then
undergoes a transition to state i*-2. In this case, it is known that the facility is in state i*-1 form
t0+n+1 to t0+n*-1 and then goes to state i*-2. However, this kind of information cannot be obtained
through inspections that take place every few time periods. Suppose that one inspection takes place
at t0+n+1 and a subsequent inspection at t0+n*. The facility will be found in state i*-1 at the first
inspection and in state i*-2 in the second. There are many sequences of holdings and transitions
that could lead to the above observation, and the event whose probability is calculated in eqn (7-27)
is just one of these. For example, suppose it is observed that a facility is in condition i*-1 at the
beginning of time period 1 and in condition i*-2 at the beginning of time period 2. The transition
from i*-1 to i*-2 can take place along three different "paths" (i.e. sequences of holdings and
transitions) corresponding to the transition taking place in period 2, 3 or 4. The probability of the
observed transition from i*-1 to i*-2 in 4 time periods is the sum of the probabilities of following
each of the three paths. The probability of following any one of the paths is obtained from an
expression such as eqn (7-27). Define:

\[ \phi_{i*-1,i*-p}(t0+n+1, t0+q | H(t0+n)) \text{ as the probability that the facility is in state i*-1 at time } \
\text{t0+n+1 and at a lower state i*-p at some later time t0+q, given the history of transitions before } \
t0+n+1. \]

This probability, called the "multi-step transition probability" is expressed in terms of the joint
probabilities of eqn (7-28) in terms of the following recursion:

\[ \phi_{i*-1,i*-p}(t0+n+1, t0+q | H(t0+n)) = \]
\[
\sum_{n^* = a^* + 2}^{q - p} h_{*, -1, *, -2}(t_0 + n + 1, t_0 + n^* \mid H(t_0 + n + 1)). \Phi_{*, -2, *, -p}(t_0 + n^*, t_0 + q \mid H(t_0 + n^*)) \tag{7-28}
\]

Equation (7-28) states that the probability of going from state \( i^*-1 \) to state \( i^*-p \) in the \((q-n-1)\) time periods between \( t_0+n+1 \) and \( t_0+q \) is equal to the probability of going from state \( i^*-1 \) to \( i^*-2 \) by time \( t_0+n^* \) and going from \( i^*-2 \) to \( i^*-p \) in the remaining \((q-n^*)\) time periods. This is summed over all possible values of \( n^* \). In other words, the probability of going from \( i^*-1 \) to \( i^*-p \) in \((q-n-1)\) time periods is obtained by summing over all possible paths along which such deterioration can take place.

The \( h(\cdot, \cdot) \) term in equation (7-28) is computed in equation (7-27) in terms of the one time period conditional probabilities of equation (7-8) and (7-9). By successively substituting for the \( \Phi \) terms on the right hand side of equation (7-28) it is possible to express the probability of any observed multi-period transition in terms of the one period conditional probabilities.

The above analysis has developed the multi-period transition probabilities for the case of a single latent variable and it's associated indicator. Extension to the multidimensional case where a \((P \times 1)\) vector of continuous variables is measured by the discrete transitions of a \((P \times 1)\) vector of indicators is straightforward, except that the algebra becomes more complicated. The derivations are not reported in this thesis.

Equation (7-28) gives the probability of occurrence of a multi-period transition that is observed in the data as the sum of products of various one step transition probabilities. The product of these probabilities is the likelihood function. The parameters of the deterioration equation (7-1a) are estimated from these probabilities using maximum likelihood estimation techniques. The link
between the expression of equation (7-28) and the parameters to be estimated is made by the choice of a particular model for the single period transition probability by making specific assumptions about the error term correlation. As described in Section 7.2, these assumptions can be used to derive expressions for the single period probabilities in terms of the parameters.

7.4 Conclusions

In this chapter, a methodology for estimating the parameters of a latent deterioration equation when the indicators are the form of discrete condition states has been developed. The models proposed here are dynamic extensions to the models formulated by McKelvey et. al. [1975] which dealt with a similar problem for a single point in time. They are also an extension to the models proposed by Heckmann [1981] which were formulations for just two states. Readers are directed to these authors for more background on models of this type.

The ultimate output of these models is a deterioration equation that predicts the propagation of underlying deterioration over time from data on observed condition states. This is the same output as is obtained from the continuous models of Chapter 3, 4 and 5. The deterioration models can then be used for maintenance resource allocation and activity selection.

In this chapter, models of deterioration were formulated, but no attempt was made to implement these on data. The data available for implementation of models of this sort is on bridge condition obtained from the National Bridge Inventory (NBI) data base. This data base contains condition ratings from over 550,000 bridges reported every year. Further details about this data are provided in [AASHTO 1980]. Unfortunately, this data is subject to many errors. [Busa, Ben-Akiva and
Buyukozturk 1981] extensively analyzed a subset of this data and found two serious problems. They are 1) omissions of maintenance history and 2) illogical data items due to possible coding or recording errors. In the light of these problems, the application of the models developed in this chapter to the data has been deferred for future research.
CHAPTER 8

CONCLUSIONS

The major contribution of the research has the development and implementation of a flexible probabilistic framework for predicting the deterioration of infrastructure facilities. The following conclusions can be derived from the research:

1) The results show that for empirical data collected from in-service pavements, there is an inherent relationship between deterioration and maintenance, since maintenance is performed in response to deterioration. With at least one model, the results showed a tendency for the highway agencies to over-compensate for the effects of deterioration. As a result, the overall effect of a 1% increase in traffic was an improvement in condition. This demonstrates clearly that the proper specification of deterioration and maintenance, therefore is as a simultaneous equation system, where the deterioration is a function of maintenance, and the amount of maintenance performed is a function of condition. In this case, failure to recognize the simultaneity would have lead to a counter-intuitive sign for the parameter estimates of the traffic variable.

b) The approach developed in this research substantially improves the fit of the deterioration model to data, by jointly estimating the performance index and the deterioration model from observed data. The infrastructure management literature has often reported that the fit to data of performance prediction models is poor. At least some of the problem of fit seems to be due to the choice of performance criteria such as the PCI or the PSI which are computed sequentially in two steps; the first step involving the calculation of a performance index and the second step being the estimation
of the deterioration model.

c) The methodology developed in this research allows the comparison, on a rigorous statistical basis, of the effect of different damage measurements on the condition of the pavement. Two latent performance characteristics were identified. The first, a roughness related characteristic, used the same measurements as the PSI, but gave greater weight to the roughness components than the PSI. The second performance characteristic was cracking or load related and was computed from measurements of rut depth, cracking extents and severities of longitudinal and alligator cracking. This is consistent with the role played by different damage measurements reported in the deterioration literature; however, until now, no rigorous method has existed in the literature for quantitatively comparing the effects of different damage measurements on condition.

d) The model with the best fit to the data is obtained when there is correlation between the error terms of the roughness related and the load related latent variables. The error terms account for unobserved characteristics of deterioration, and the correlation between the error terms indicates that the underlying characteristics that affect deterioration influence the propagation of both roughness and cracking. The latent variable technique offers the flexibility of including a wide variety of such interactions.

e) The percentage change in condition as the result of a unit change in the explanatory variables (for example, an increase in age of one year, or a 1% increase in traffic) is different for the different performance measures. This indicates that the sensitivity of the condition to changes in explanatory variables varies according to the performance measure used. Maintenance strategies selected on
the basis of the latent variable models will therefore be different from strategies selected using the PSI. An important area for future research is to study the effect of the maintenance strategies predicted by the use of different performance measures.

f) The question of how to incorporate user costs into a latent variable model and how to express these costs as a function of latent condition is another important research issue. In order to compare between different maintenance alternatives it is necessary to be able to accurately predict the costs associated with the pattern of deterioration caused by a particular maintenance activity. Methods for simultaneously estimating a cost equation along with the deterioration and maintenance models should be investigated.

Summarizing, the latent variable method produces results that fit the data better but are different from the results obtained from traditional methods. The following points of investigation are necessary to obtain a more complete understanding of the results before the method is ready for practical use.

1) **Non linear models**: All the specifications estimated so far have been linear in the parameters. The reason for trying the linear specification is that it was a good starting point and was easy to estimate. A linear deterioration model might be adequate for predicting the deterioration of pavements in the middle of their lives, but empirical evidence has shown that for old and new pavements, condition is a non linear function of time. Extensions to the model should include some non linear specifications as well.

2) **Time series data**: The time series information contained in the data set used for the case study
could not be utilized because it was not possible to distinguish the sequence of maintenance and inspections over time. As a result, the entire data was averaged over the five year period. The model therefore could not incorporate temporal correlations in the error terms of the deterioration and measurement models. These correlations are likely to be important in the specification of a realistic model that predicts the deterioration of the pavement over time, and need to be investigated in future specifications with other data sets.

3) Consequences for strategy selection: As mentioned previously, the pattern of deterioration predicted by the latent variable approach is different from that predicted by traditional performance measures such as the PSI. An extension to the current research on deterioration models is the development of a realistic probabilistic strategy selection model that can compare between the maintenance strategies selected by the use of different performance measures. The comparison will be based on the implementability of the predicted strategies, and the costs associated with the implementation.

4) Inclusion of costs into the model framework: One of the critical requirements for a model that compares between different maintenance alternatives is the ability to accurately calculate the cost to the user and to the agency from the pavement being in a particular condition. The agency cost is relatively easy to measure, since it is directly connected to the extent of maintenance performed. The user cost, expressed as a function of pavement condition, is difficult to estimate when the condition measure is latent. One option is to treat the "true" user cost as a latent variable. This latent user cost is a function of condition, also latent. The indicators for user cost are the different measures that traditionally define out of pocket costs to the user; e.g. fuel costs, maintenance and repair costs, costs associated with equipment wear and tear etc. Other "cost" measures that may be
used are comfort ratings calibrated on some arbitrary scale (for example, 10=most comfortable, 1 = least comfortable). If \( C \) stands for a vector of latent user cost variables, and \( F \) for a vector of measured costs, then the following system of equations can be formulated as an extension of equation (3-3):

\[
S = g_1(A, X_1; \beta) + \eta_1 \quad (8-1a)
\]

\[
A = g_2(S, X_2; \gamma, \Gamma) + \eta_2 \quad (8-1b)
\]

\[
l = h(S; \lambda) + \epsilon_1 \quad (8-1c)
\]

\[
C = k(S) + \eta_3 \quad (8-1d)
\]

\[
F = m(C) + \epsilon_2 \quad (8-1e)
\]

Equation (8-1d) represents the latent cost as a function of latent condition. Equation (8-1e) is the measurement model for the user costs that expresses the measured cost as a function of the latent cost \( C \). In this specification, the connection between latent cost and the associated indicators is not established directly; the connection arises through the fact that both the indicators and the latent costs are functions of latent condition.

The system of equations (8-1) is one method of including user costs in a latent variable deterioration model. Further research is required to determine whether such a model can be estimated. This is an important issue for future research, since the ability to incorporate cost information into a deterioration model is necessary in order to be able to compare between the effect of different maintenance activities.
REFERENCES


White, H. "A heteroskedasticity consistent covariance estimator and direct tests for heteroskedasticity", *Econometrica*, 48:817-838