THE EVOLUTION OF TRANSPORT SYSTEMS:
AN ANALYSIS OF TIME-STAGED INVESTMENT
STRATEGIES UNDER UNCERTAINTY

by

WAYNE M. PECKNOLD

B.S., Michigan State University
(1963)

S.M., Massachusetts Institute of Technology
(1965)

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

at the

Massachusetts Institute of Technology
September 1970

Signature of Author
Department of Civil Engineering, September 8, 1970

Certified by
Thesis Supervisor

Accepted by
Chairman, Departmental Committee on Graduate Students

Archives
OCT 7 1970
ABSTRACT

THE EVOLUTION OF TRANSPORT SYSTEMS: AN ANALYSIS OF TIME-STAGED INVESTMENT STRATEGIES UNDER UNCERTAINTY

by

Wayne Martin Pecknold

Submitted to the Department of Civil Engineering on September 8, 1970 in partial fulfillment of the requirements for the degree of Doctor of Philosophy

The primary purpose of this study is to explore the problem of making time-staged transportation investment decisions in the face of uncertainty. The transportation investment problem is presented in terms of a time-related sequential investment framework. This framework recognizes first of all, that investments are usually implemented as a series of staged sequential increments to a fairly extensive existing system, and second, that there is substantial uncertainty over the future demands for those investment increments.

A basic stochastic sequential time-staging model is developed which is capable of handling supply-demand interdependencies, network connectedness, budget constraints, and longer-run activity system dependencies on transport investments. The model is used in a first set of experiments to study the tradeoffs between opportunity losses from excess capacity and scale economy losses from staging, under assumptions of deterministic, static and dynamic uncertainty conditions for the problem of project design. Attention is also devoted to the more general philosophical question of searching for flexible transportation systems which can adapt to a wide range of uncertain future demand conditions.

The use of a descriptive, non-analytic simulation model for transport flows which recognizes both uncertainty and the multistage nature of investment alternatives results in an extensive form multistage decision tree of extreme dimensions. Approximating procedures, called pruning rules and terminal evaluation functions, are developed to heuristically reduce the solution space and make application of the sequential decision model in extensive form feasible for large network problems.

Finally, the general model is extended to include a Bayesian updating procedure to revise probabilities at each stage of the tree. This adaptive version of the sequential decision model produces a
series of conditional decisions or strategies which are then responsive to both uncertainty of demand and the true sequential nature of implementation.

To accomplish these objectives, both the model and the approximating procedures are implemented as a series of computer programs (DECISN) which are then used to conduct some initial experiments to explore the concept of the sequential decision model as applied to transport investment planning. The programs both simulate the supply-demand network flow patterns at a single point in time as well as produce shifts in demand over the longer-run for conditions of a stochastic demand formulation.

Briefly, conclusions based on experiments with the model indicate investment patterns can and do change for specific patterns of uncertain demand. Substantial savings are realized by considering investments in a staged sequential investment framework. Secondly, the general framework can be applied to problems of a small scale without incorporating the pruning rules and terminal functions. For large project design problems or even modest network problems, the search space generally requires these functions. Results from using specific pruning rules and terminal functions indicate that good heuristic investment patterns can be developed without searching the entire tree. Finally, an adaptive mechanism is necessary when there is additional uncertainty over the underlying demand parameters. Its use produces investment strategies which are both responsive to changing demand conditions and conditional on previous investment sequences.

Thesis Supervisor: Marvin L. Manheim
Title: Associate Professor of Civil Engineering
ACKNOWLEDGEMENTS

The research reported in this dissertation was made possible through support provided by the U.S. Department of Transportation, Office of Systems Analysis, through contract 7-351 40 (DSR 70386). Additional support for portions of this research were provided by the M.I.T. Inter-American Program in Civil Engineering and by the Urban Systems Laboratory of M.I.T. The author also gratefully acknowledges the support of the M.I.T. Urban Systems Lab time-shared computer facility.

The author would like to extend his appreciation to Professor Marvin L. Manheim, associate professor of Civil Engineering at M.I.T., for his supervision of this thesis, for his interest and most helpful encouragement at various points in this work and during the author's career. His pioneering efforts in the field of transport systems analysis, his approach to some of the issues of transportation and its relation to the larger problem of social consequences has affected to a large degree my thinking and my philosophy in approaching problems.

The author would also like to thank a number of individuals for the many discussions and comments at various points in this effort. In particular, I would like to thank William F. Johnson for the many hours spent discussing problems of thesis research; Seigfried M. Breuning for his interest and encouragement in the formative stages of this research and the author's career; Professor Paul O. Roberts, Jr., currently of the Harvard Business School, whose research, philosophy and only what can now be called classical work in the development of transport planning models has greatly stimulated the author at various points in the last few years; Earl R. Ruiter for his reading of a number of drafts of this research.

In addition, the author also wishes to express his appreciation to Jon Bottom for his long hours spent programming and running experiments and for his interest in the research itself; Len Sherman, for his help in programming parts of this research; and to Russel Fiddler, who provided the illustrations.

For her assistance in solving some of the many problems associated with producing a document of this size, the author wishes also to thank Mrs. Nora Scanlan. For her typing of portions of the final document under extreme time constraints, a great deal of appreciation is extended to Mrs. Sue Sager.

Finally, the author would like to express his deepest appreciation to those who endured the conditions of thesis research - my family - and especially to one who typed a number of the early and sometimes cryptic drafts and a significant portion of the final document, and most importantly, who offered encouragement and enthusiasm at points when they were very obviously most needed - I am extremely thankful to my wife Jeannie.
<table>
<thead>
<tr>
<th>TABLE OF CONTENTS</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>CHAPTER 1.0 INTRODUCTION</td>
<td>11</td>
</tr>
<tr>
<td>1.1 Overview of the Problem</td>
<td>15</td>
</tr>
<tr>
<td>1.2 Purpose and Assumptions</td>
<td>37</td>
</tr>
<tr>
<td>1.3 Structure of the Study</td>
<td>42</td>
</tr>
<tr>
<td>CHAPTER 2.0 THE TRANSPORT INVESTMENT PROBLEM</td>
<td>45</td>
</tr>
<tr>
<td>2.1 The Nature of Transport Investment Planning</td>
<td>47</td>
</tr>
<tr>
<td>2.1.1 The Objectives of Transport Investment Policies</td>
<td>47</td>
</tr>
<tr>
<td>2.1.2 The General Planning Methodology of Transport Investment Planning</td>
<td>51</td>
</tr>
<tr>
<td>2.1.3 Scope of this Study</td>
<td>61</td>
</tr>
<tr>
<td>2.2 Statement of the Problem</td>
<td>63</td>
</tr>
<tr>
<td>2.2.1 The Issues of Project Design and Analysis</td>
<td>64</td>
</tr>
<tr>
<td>2.2.2 Extension to the Network or &quot;Program Selection&quot; Problem</td>
<td>85</td>
</tr>
<tr>
<td>2.2.3 Uncertainty</td>
<td>88</td>
</tr>
<tr>
<td>2.3 The Sequential Decision Framework</td>
<td>91</td>
</tr>
<tr>
<td>2.3.1 Definitions and Terminology</td>
<td>91</td>
</tr>
<tr>
<td>2.3.2 The Definition of Strategies and Sequential Decisions in an Adaptive Sense</td>
<td>98</td>
</tr>
<tr>
<td>2.4 Survey of the Literature</td>
<td>101</td>
</tr>
<tr>
<td>2.5 Summary and Conclusions</td>
<td>113</td>
</tr>
</tbody>
</table>
CHAPTER 3.0 POSSIBLE TECHNIQUES FOR THE
SEQUENTIAL DECISION MODEL................................. 116

3.1 Description of the Problem................................. 117

3.2 Mathematical Formulation of the Problem................... 121
   3.2.1 Link Decision Variables........................... 121
   3.2.2 Investment Programs............................... 121
   3.2.3 Link Investment Matrix............................ 124
   3.2.4 State Transformation Matrix........................ 126
   3.2.5 Equilibrium Flows at Any Stage...................... 128
   3.2.6 The Objective Function............................ 130
   3.2.7 The Combinatorial Effects of
   Time-Staged Investment Sequences....................... 132

3.3 Techniques for Solution (Alternative
Search Procedures).............................................. 137
   3.3.1 Indirect Mathematical Programming
Search Procedures........................................ 138
   3.3.2 Direct Search Procedures........................... 149

3.4 A Decision-Tree Procedure for
Transport Investments........................................ 151
   3.4.1 Logic of the Tree-Tracing Procedure................ 152
   3.4.2 A Formal Statement of the Algorithm............... 155
   3.4.3 A Set of Heuristic Approximations.................. 161

3.5 A Comparative Analysis of Search
Efficiencies.................................................. 162

3.6 Summary and Conclusions.................................. 167

CHAPTER 4.0 A PROPOSED HEURISTIC SOLUTION PROCEDURE
FOR THE SEQUENTIAL DECISION FRAMEWORK..................... 169

4.1 Description of the Approach................................ 170

4.2 The Proposed Heuristic Search Procedure................... 174
4.2.1 Determining the Cut-Off Point Using Explicit Pruning Rules .......................... 174

4.2.2 Adjusting for Variable Pseudo-Horizon Cut-Offs using Terminal Evaluation Functions ................................................. 179

4.2.2.1 A Collapsed Stage Space ................................................. 182

4.2.2.2 The Myopic Rule ......................................................... 183

4.2.2.3 A Reduced State Variable Approach ................................. 186

4.2.2.4 Approximating the Alternative Space at a Terminal Point ............. 188

4.3 The Adjusted Tree-Climbing Algorithm ........................................... 192

4.4 Application of the Procedure to the Transport Investment Case ................. 198

4.4.1 Uniqueness of Transportation: Irreversibility and Longevity ................. 198

4.4.2 Selection of a Horizon ..................................................... 199

4.4.3 Adjusting Terminal Conditions for Distorting Edge Effects .................. 201

4.5 Summary of the Search Procedure ............................................... 204

CHAPTER 5.0 COMPUTATIONAL EXPERIENCE WITH THE PROPOSED SEQUENTIAL MODEL .......................... 207

5.1 Introduction: General Description of the Problem .................................. 209

5.2 Structure of the Model .......................................................... 212

5.2.1 Supply Functions ........................................................... 212

5.2.2 Aggregate Supply: Network Effects for the Project .......................... 215

5.2.3 The Demand Function ...................................................... 217

5.2.4 The Socio-Economic Activity Shift Allocation Model ....................... 220

5.2.5 Equilibrium at a Stage and Net Benefits .................................. 226
5.3 Optimal Capacity Expansion under Conditions of Assumed Certainty and Uncertainty

5.3.1 Comparison of Absolute versus Comparative Advantage

5.3.2 Trade-offs between Staging and Scale Economies for Deterministic Demands

5.3.3 Flexibility, Adaptability, and the Staging Problem under Uncertainty

5.3.4 Comparison of Investment Sequences for Deterministic and Uncertain Demand Parameters

5.3.5 The Effect of a Sequential vs a Non-Sequential Decision Procedure

5.3.5.1 Results for the Case of Long-Run Supply-Demand Independence

5.3.5.2 Incorporating the Accessibility-Related Activity Shift Model

5.3.6 The Sensitivity of Investment Policies to Changed Initial Conditions

5.4 Experimental Results with the Proposed Approximation Procedures

5.4.1 Horizon Approximation

5.4.2 Results Using the Myopic Search Rule

5.4.3 Results of Stage Aggregation

5.4.4 Effects of Variations in the Number of Classes and Selected Probability Spines

5.4.5 Comparison of Myopic and Stage Aggregation as Terminal Evaluation Functions

5.5 Comparison of Computation Times for Variations in Problem Size

5.6 Summary and Conclusions
# Chapter 6.0 An Adaptive Sequential Model for Capacity Expansion Under Uncertainty

## 6.1 Restatement of the Demand Uncertainty

## 6.2 Development of a Learning Adaptive Model for Capacity Expansion under Uncertainty

### 6.2.1 Introduction

### 6.2.2 Mathematical Formulation of the Bayesian Approach for the Multistage Sequential Investment Problem

## 6.3 Incorporating the Adaptive Mechanism into the Basic Multi-Period Algorithm

### 6.3.1 A Discrete Version: DECISN with Bayesian Revision

### 6.3.2 Variations in the Subjective Stage Probabilities for Different Priors and Conditionals

## 6.4 Computational Experience with the Learning Adaptive Model for the Project Design Capacity Expansion Example

### 6.4.1 Computational Results

### 6.4.2 The Effect of the Learning Adaptive Model on Technology Adaptation

### 6.4.3 Effects of Variations in the Conditional Learning Function and Prior Distribution

## 6.5 Conclusions and Extensions

## Chapter 7.0 Summary and Conclusions

### 7.1 Summary of the Research Effort

### 7.2 Conclusions

### 7.3 Extensions and Directions for Future Research
<table>
<thead>
<tr>
<th>REFERENCES</th>
<th>348</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIOGRAPHICAL NOTE</td>
<td>372</td>
</tr>
<tr>
<td>APPENDIX A NOTATION AND SYMBOLS</td>
<td>375</td>
</tr>
<tr>
<td>APPENDIX B LIST OF FIGURES</td>
<td>380</td>
</tr>
<tr>
<td>APPENDIX C LIST OF TABLES</td>
<td>383</td>
</tr>
<tr>
<td>APPENDIX D AN EXTENSION TO THE LEARNING ADAPTIVE SEQUENTIAL MODEL FOR UNKNOWN CONDITIONALS</td>
<td>D-1</td>
</tr>
<tr>
<td>D.1 LEARNER I: An Adaptive Model for Unknown Conditionals</td>
<td>D-1</td>
</tr>
<tr>
<td>D.1.1 Introduction</td>
<td>D-1</td>
</tr>
<tr>
<td>D.1.2 Method</td>
<td>D-3</td>
</tr>
<tr>
<td>D.1.3 A Comment on Initial Assumptions</td>
<td>D-9</td>
</tr>
<tr>
<td>D.2 LEARNER II: An Exact Formulation</td>
<td>D-10</td>
</tr>
<tr>
<td>D.2.1 Introduction</td>
<td>D-10</td>
</tr>
<tr>
<td>D.2.2 Method</td>
<td>D-10</td>
</tr>
<tr>
<td>D.3 Notes to Appendix D</td>
<td>D-13</td>
</tr>
</tbody>
</table>
Chapter 1

INTRODUCTION
Chapter 1
INTRODUCTION

The present congestion found in many large urban cities and some regional corridors is a major problem currently facing the city, state, and federal transportation agencies of this country. The problem of alleviating this congestion, and in general, providing adequate transportation service for both public and private interests is not simply one of locating the congestion and building enough capacity to adequately relieve it, however. At the technical level, there are serious questions that must be answered such as how much capacity should be provided, which are the most seriously congested facilities, and should we merely expand the existing facility or would a new or different mode be more appropriate, and so on. At a more goal oriented level, there are even questions as to the appropriate measures of benefits, the incidence of both benefits and costs on users and non-users alike, and the equity of the investment decision.

More importantly, facilities are generally provided not only to relieve existing congestion but to also meet future expected increases in demand. Investments provided now, because of long economic lives, and because they can constrain subsequent investment alternatives, have considerable future impacts associated with them. Because of this and because funds available for investment are obviously limited, decisions concerning investments must not only consider where and how much capacity to provide but also when should it be provided.

This task of planning transport investments is part of a broad class of problems generally referred to as the problem of the
allocation of scarce resources; this allocation usually occurs over both space and time, and in some instances, to systems which have fairly significant long range future consequences. In the transport case, it is exemplified by the charge given by the White House to the Department of Commerce for the Northeast Corridor Transportation Project, which was to develop a regional transportation plan through the year 1980 for the region known as the Northeast Corridor [1].

In the past, because of constraints of both time and money, there have been a number of simplifying assumptions usually made in approaching this task. The early urban planning studies have typically involved only what can be called an extremely limited and cursory examination of a few investment plans for some future and rather arbitrary design year. Moreover, this analysis has commonly been made using the following assumptions:

(1) demand is assumed to be insensitive to the level of transportation service provided, i.e., demands are inelastic and unaffected by alternative plans;

(2) technology is relatively stable and non-changing over time; and

(3) the future is known and can be predicted with certainty.¹

Experience over the last decade in both the urban and regional areas has proven that these assumptions are no longer tenable,

¹See, for example, either of the two earliest large urban transportation planning studies [2,3].
except perhaps in some very extreme cases. Demand has been shown to be extremely dependent both on the socio-economic characteristics of the region, and on the levels of service, in terms of travel time, cost, frequency of service and so on, offered by the transport system [4]. Transportation technology is anything but non-changing as new forms of transportation are rapidly being developed; for example, VSTOL in the regional context [5], Dial-a-bus and dual-mode vehicles in the urban environment [6,7]. And, finally, the assumption of deterministic future conditions is clearly not what has been observed to be happening in both the regional and urban areas over long periods of time; activities have shifted substantially from the patterns that were expected to occur 10 or 20 years ago [8].

Accordingly, some of the more recent transportation studies have seriously questioned the first two of these assumptions and have adjusted their analysis procedures to include both new systems, such as high speed rail, VSTOL, etc., and some fairly sophisticated supply-sensitive demand functions in their prediction and evaluation of the consequences of future plans [9,10]. The third assumption, however, remains as a relatively unexplored area of research in transport planning. Little if anything has been done to incorporate the effects of a continually changing and highly uncertain future environment.

The purpose of this thesis will be to explore the problem of making investment decisions for transport facilities when faced with highly changing and uncertain future conditions. This problem is generally referred to in the literature as "the investment problem
under uncertainty" [11,12,13,14,15]. The study will approach the transport investment decision as a time-related sequential investment problem, recognizing the fact that investments are usually implemented as a series of staged sequential increments to a fairly extensive existing structure. More specifically, we will explore the application of the general stochastic sequential decision model to the transportation investment problem, concerning ourselves with both approximation procedures which make the approach feasible, and the more general philosophical question of searching for flexible transportation systems which can adapt to a wide range of uncertain future demand conditions.

This chapter will give a brief overview of the transport investment problem, the nature of the sequential investment model as applied to transportation, including the assumptions necessary to reduce the problem to a manageable size, and the general structure of the study.

1.1 Overview of the Problem

Because of increasing congestion, pollution, and in general, a dissatisfaction with the service offered by existing transportation systems, it is becoming apparent that transportation is fast becoming one of the major problems of our environment. Interest in one or more of these problems has produced innumerable suggestions over the last 10 years or so for improving these existing systems. Proposals have ranged from the traditional, but somewhat dated, more "balanced" transportation schemes [16] for getting people out of their cars and
into mass transit systems, to some very serious and specific concerted
efforts to improve on our present systems. For example, in urban areas
there are a number of promising new proposals; they range from such
short-range technologies as the Dial-a-Ride system of computer-aided
door-to-door mini buses [17], to the much longer-range concepts of
automated dual-mode vehicles which will operate at much increased ca-
pacities and require less than one-tenth the space of existing freewe-
ways [18].

There have been similar reactions in the regional transpor-
tation context. Airport terminal and access congestion, schedule
delays and delays in intercity travel time have resulted in a number
of suggested improvements. These too have ranged from the seemingly
simple solutions, such as expanding existing capacity, to the more
exotic solutions of constructing completely new modes of transportation.

Interest in transportation during this period has not been
limited to new technology and hardware changes alone, however. There
has been a substantial accompanying interest shown for the softer side
of transportation—for the techniques and methodology of analysis and
evaluation of alternative investments. One of the most significant
conclusions to evolve out of the sixties by those concerned with long-
range planning was that the traditional analysis techniques of trans-
port investment planning were not fully capable of providing answers

---

1 Short-range in the sense of expected implementation time. Dial-a-
ride has been undergoing concentrated development and is soon to
undergo extensive testing in the U.S. as a demonstration project
[19].
to the when, where, and how much of new capacity to add, whether it be an existing or new technology. Therefore, perhaps even more significant than the suggested hardware changes has been the way in which the field of transportation planning has evolved in its approach to the problem of planning expansions and additions to existing transportation systems. For a number of reasons, the "problems" of transportation have broadened considerably in scope during this period.

For example, it is now generally recognized that investments in transportation have a much greater impact on our environment than simply adding to, or decreasing the more obvious problems of pollution, congestion, and so on. This impact is the result of transportation's strong interdependence with other sectors of the economy. As stated by the Harvard Transport Research Program [20]:

"The transportation sector of the economy is closely linked to the other sectors and change within the transport sector has obvious repercussions throughout the entire economy."

In the case of developing countries, transportation can often be crucial to the achievement of economic growth and a country's development plan. Recognizing this fact, many developing countries have committed a substantial portion of their budgets to the transport sector.

"Of ten countries where information for a recent year is readily available, investment in transport has ranged from around 2 percent to 5 percent of Gross National Product and from 10 to nearly 50 percent of total fixed investment is borne by governments. In several countries, transport has accounted for over 25 percent of total public investment." [21]
For urban and regional areas of developed countries, although to a lesser degree, there is a similar influence of transportation on the economic sector. The presence, or lack of transportation facilities has, in some cases, significantly affected the locational patterns of industrial and commercial development,\(^1\) as well as determined the future patterns of residential growth. In particular, in the urban areas during the past decade or so, the automobile has been conspicuously dominant in determining urban travel patterns; so much so that in some cases the very quality of urban life and metropolitan development can in some sense be said to be determined for some years to come.\(^2\)

In addition to the dependency between the transport sector and other sectors of the economy, it has also been generally recognized...

\(^1\)Route 128 in Boston and the shift of industry from the core area is a classic example of the effect of transportation on land use patterns [22].

\(^2\)As many writers point out, however, the problems of economic development and the problems of metropolitan growth are not solely problems of transportation development and we must be careful not to demand too much from transportation in the way of social improvements. For example, as Meyer, Kain, and Wohl [23] have stated it:

"The sheer variety of needed urban services reflects the fact that the sources of urban problems are much more complex than, let us say, a clear-cut inadequacy of transportation facilities."

Nonetheless, even if a lack of transportation facilities cannot be solely blamed for some of the more pressing problems of the environment, it is clear that transportation can be an extremely powerful catalyst in alleviating some of these problems.
that there are severe dependencies within the transport sector itself that seriously complicate the investment decision problem. Some of the more important of these dependencies are the following:

A. Network or Project Interdependencies

Links in a network are topologically connected. Therefore, the benefits and costs attributable to an improvement in one part of the network can be affected in some way by the nature of the rest of the network, i.e., a change in the performance characteristics of a specific link or in the connectivity of the network, can affect the volumes and therefore the costs and benefits of other links in the network.

For example, in the network of Figure 1-1, improving link 4-2 will improve the service to all those interzonal flows using this link. This improvement may also induce interzonal flows between zones (represented by the shaded arrows) not using this link to shift to a path that does contain this link. This produces two effects. First, it increases the volume on link 4-2 and can perhaps decrease its level of service in terms of speed, or operating cost; and secondly, it also improves the service level on all links formerly used by the switching volumes. These effects tend to induce travellers using still different paths to consider using these secondary paths, which changes the

---

1Although almost all current large-scale planning studies have incorporated dependencies of this type, we have included this category for completeness since traditional project analysis approaches in the pre-urban planning process period, and even some current study approaches, have been known to ignore it. We will discuss this further in Chapter 2.0.
--- Link Additions

Figure 1-1
The Relationship of Project Design to the Overall Network Structure
levels of service on these paths, which induces..., and so on. Thus, in general, a change to one link in a transportation network can have effects on every other link in the network, merely by the nature of a network's topological interconnectedness.

B. Modal Dependencies

Investment in a particular mode also has implications for both competing (or substitutable) and complementary modes, (e.g., investments in highways have effects on transit demand; investment in air facilities have effects on rail and intercity highway demand). Planning modal investments without recognizing these dependencies can sometimes lead to severe distortions in the final investment pattern and differences in the other sectors of the economy. Continuing with the quote from the Harvard report reference [24]:

"...changes made to one of the modes of a transport system can induce changes in all others since physical differences can produce cost differences which have implications for prices, profits, choice of mode, flow volumes, utilization rates and market areas to name a few."

C. Supply-Demand Dependencies

Aside from the simple "sharing" and shifting of a fixed demand between modes and links that we have just described, there is an additional more complex interdependency based on the economic theory of supply and demand. Applied to transportation, it implies that the total demand for travel will also depend in part on the characteristics of the supply of transportation offered. In general, the more desirable the offered level of service, the more demand there will be for the facility. From the point of view of the complete network, however,
the demand for travel is not just a function of the presence of any particular link, but depends in part on the complete path supply of capacity (and alternate paths) of a number of links from the initial trip origin to a final destination. Consequently, demand between any two modes in the network can then either be inhibited or encouraged by the availability of both specific links and the total network "supply" of transportation between these modes. In terms of Figure 1-1, improving link 4-2 can induce travellers from all origins who were not previously making a trip to use paths containing this link, in addition to simply causing shifts between alternate paths. This, in turn, again has repercussions throughout the entire network.

D. Financial Dependencies

The funds available for transportation investment are generally limited by budget constraints. Transportation is merely one sector from among many, and resources are obviously constrained by the currently available funding mechanisms, and at a larger scale, by the resources of the country. Within the transport sector itself, generally there are also limits on the funds that can be expended for any one mode in any given year. Constraints on available capital, therefore, imply that a financial dependency exists between projects since capital expended on one project cannot be spent on another.

E. Timing Interdependencies

Finally, decisions to invest in transportation facilities must consider the dimension of time as well. Initial decisions concerning desirable transportation systems have implications for future investment alternatives. Moreover, there is usually an extensive
structure to the existing transportation system which provides an inducement to continue adding incrementally to this system; to vehicle numbers, to track or lanes, or to new links in the network. A shift in direction to a totally new network (and even more so, to a new technology) is not likely to be easily implemented because of the tremendous inertia present in the existing system. Thus, the fact that fixed facility investments can be constructed in different sequences, coupled with their dependence on the other sectors of the economy, implies that there can be severe timing interdependencies as well, and

"...if true timing interdependencies exist, where the order in which projects are undertaken changes costs and/or benefits of the projects, the problem becomes considerably more complex." [25]

Thus, it is clear that the problem of planning transportation improvements is indeed a complicated one. ¹

Historical Approaches to Investment Planning

Early approaches to transport investment planning have explicitly recognized some of these dependencies and understandably ignored others for a number of reasons. For example, the earliest of the urban

¹These preliminary statements are not meant to be a complete characterization of the transport planning problem, but simply an indication of the complexity of the problem and how transport planning has come to recognize the problems of investment analysis over the past decade. For a more complete discussion of the complexity of the problem of planning transport investments, see Manheim et al [26], Meyer and Straszheim [27], or Fromm [28].
transportation planning studies largely ignored both modal and sectoral dependencies for two reasons: a) there was essentially only one mode which dominated all others (i.e., highways), and b) sectoral dependencies were at best something of a mystery to transport planners.

The primary concern of these early studies was with a fairly static and deterministic analysis of comparing alternative "master plans" [29,30] for some particular future design year. These studies also generally assumed that the technological alternatives were constrained to the available technology at that time. Additionally, demand for transport service was assumed fixed and independent of the level of service offered. Most of these early urban studies first considered a few alternative plans and then searched for the least cost system of the group which would meet the predicted (independent) demands in the target year, some 20 to 30 years in the future.

More recent transport planning approaches have recognized the implications of ignoring modal, sectoral, and supply-demand dependencies and have developed a rather sophisticated methodology of planning models for evaluating network investments. (See, for example, [31] or [32].) As the primary example in the urban studies, considerable effort has been directed at capturing sectoral dependencies--the prediction of residential and industrial effects due to new transport investments. This effort has been called land-use modelling in the urban area studies [33,34,35]. (We will use the term activity shift models in later discussions.)

The purpose of these models is to be able to predict the population, residential and industrial changes, and in general, the
effects on the economy caused by changes in the transport structure. The more elaborate of these approaches to date have in fact, devoted as much effort to the details of modelling the economic sector as to the simulation of the more direct, transport-related costs and benefits. The most notable of these are the larger, more recent urban planning studies [36,37], the Northeast Corridor Study [38], and the Harvard-Brookings study for developing countries [39].

The use of an explicit time-staged multi-period approach has also been recognized by most of these more recent planning approaches which we will discuss in more detail in Chapter 2.0. However, the design of these systems of models has been directed towards solution for either (or both) small numbers of alternative plans or deterministic demand conditions, both of which are not typical of the real world.

**Uncertainty, Flexibility and Sequential Decisions**

There has been sufficient interest and development in transport planning that most of the previously described dependencies can now be handled in one form or another. There are still at least two major problems with the methodology of current planning approaches, however. The first is a direct result of the fact that the environment with which the transport system must interact is very rarely as predictable as assumed, no matter how sophisticated the planning approach. There is considerable uncertainty over a) the technological options available for investment in terms of feasibility and performance characteristics, and b) more importantly, the future demands for the transport services provided by new, and in many cases, even existing
technology. Current approaches to transport planning have all generally failed to recognize this uncertainty in predicting the future consequences of transport investments. As emphasized by Marglin [40]

"By definition, present investment yields its returns only as the future unfolds, and the results of all investment are therefore inherently uncertain. Since public sector projects tend typically to be more durable than private projects, uncertainty is consequently all the more important in the economic analysis of the public investment."

The second major problem with current approaches to planning is related to the fact that transportation facilities are rarely implemented as one massive investment in a single time period. More generally, they are implemented as a series of staged sequential additions to an extensive existing system. Each stage can take a number of years to complete. Concurrently, however, the environment continues to change; land use patterns, housing markets, and travel patterns will have shifted and adjusted to a myriad of outside forces. Using the target year planning approach described previously (with or without subsequent staging of that target year plan) still fails to capture the true sequential nature of investment implementation. As described by Manheim, et al [41]:

"In general, the decision-maker not only has the option of immediate actions—particularly transportation system changes—but also of deferring implementation of an action in order to acquire more information about the problem."

Thus, there is usually sufficient time to reconsider investments—to revise decisions if outcomes do not occur as anticipated.

Unfortunately, there have been very few studies that have undertaken to explore the effects of uncertainty and sequential
implementation procedures on the investment policies of transportation planning agencies. Fewer yet have attempted to actually incorporate it in the design of alternatives for evaluation in planning models.

The solution to the problems caused by a changing and highly uncertain future is relatively straightforward in principle. It simply involves explicitly recognizing that fact in both the technological decisions we make and in the planning models themselves. There are several ways in which this can be done. On the technology side, if demands are fairly uncertain (as is the case with most new technologies) demonstration projects can help to reduce the uncertainty of market response to new systems as some recent studies have suggested [42,43, 44].

In principle, this approach is similar to a staged implementation policy for the case of existing system expansion. For example, in the case of a new highway facility designed to serve a region or a corridor where full capacity will not be needed for some time beyond the opening of that facility, it may be desirable to consider a two-stage construction sequence of the facility instead of providing the full capacity all at once. This has the dual effect of reducing the losses from as yet unneeded capacity (which can be used elsewhere in other investment programs), and of reducing the risk of loss if demand does not in fact materialize. On the negative side, there is a possibility of losing scale economy effects which are a positive derivative of many large-scale, durable investment programs. These tradeoffs are the fundamental issues to be considered when decisions involve long-
range durable investment programs which have substantial future consequences.

A second technological method for accommodating uncertainty is to directly and explicitly build adaptability into the transport system itself. The substitution of a less capital-intensive technology for a high fixed cost system can to some degree reduce a certain amount of "overcommitment" to a mode. For example, the use of inexpensive bus systems which use existing city streets involves a much smaller commitment than a fixed right-of-way and capital intensive rail rapid transit system would entail. A system that is highly adaptive can in effect "shift" resources to where and when demand materializes. On the negative side, the loss again is likely to involve lost scale economies that result from durable systems, in addition to the higher expected operating costs that generally accompany less durable systems. The degree of adaptability required will, of course, depend on the specific costs of alternative systems and on the degree of uncertainty over the magnitude and spatial pattern of demand.

The very opposite approach, i.e., incorporating more fixed cost, can be used in a similar manner to provide a system which is able to handle a wide range of volumes and, while it is inefficient for any one output volume, it is relatively efficient for this range of outcomes, i.e., it has an inherent flexibility associated with it. For example, in terms of the unit costs of alternative technologies shown in Figure 1-2, technology V is said to be more flexible than any of the other alternatives, since it is passably efficient for a wide
Figure 1-2
The Flexibility of Technology
range of outcomes, even though it is not optimal for any particular volume level.

In terms of network investments, the adaptability\(^1\) of an initial investment can be measured in some sense by the number of good alternative future networks which it may evolve to. For example, if there is considerable uncertainty about future outcomes, an initial investment pattern which can evolve to a number of end-states which are passably efficient for a wide range of conditions is considerably more desirable than the otherwise "best" initial investment which can reach only a few plans and is inefficient for extreme outcomes. Figure 1-3 represents a number of potential time-staged investment sequences; each branch of the tree represents a different set of investments which have been added to the network at a previous stage. The directional branches emanating from any network at a stage indicate which plans in the next stage are feasible. A simple comparison of alternative staged sequences will determine which sequence is optimal and we proceed to construct the initial stage of that sequence. This is accomplished simply by comparing alternative paths in the tree. At the conclusion of the initial period of investment, however, things may have

---

\(^1\)The terms flexibility and adaptability were first used by Stigler [45] in 1939, discussed by Cole [46] and numerous other authors subsequently. The terms are still rather ill-defined; however, it is certain that they are not free, i.e., one must pay to incorporate flexibility or adaptability (and there will be a tradeoff between them), but in the face of uncertainty and the world of expected values, the overall loss (expected) may well be less for these kinds of designs. We shall return to these terms in subsequent sections of this study.
Figure 1-3
The Flexibility of Initial Network Programs
changed so much that the second stage investment may either be hopelessly underdesigned or completely unnecessary. Therefore, the sequence of subsequent staged investments may have to be revised to account for these changed conditions. If our initial first stage decision reflected the possibility of changing conditions in the first place, revision may be fairly straightforward and relatively costless. If it ignored the possibility of second stage revision and a changing environment, it may either involve substantial additional capital outlays or have locked us in on an investment sequence which will be difficult to alter.

In terms of this figure, plan adaptability, therefore, should measure the ability of an initial investment pattern to evolve to a number of good plans (see Gupta and Rosenhead [47]). For example, initial investment plan \( \Lambda_{31} \) is capable of reaching any of the proposed end-states, whereas \( \Lambda_{11} \) and \( \Lambda_{21} \) are much more constrained in their potential final plans. Therefore, it is a much more adaptable initial investment.

In order to consider adaptability and revised decisions, however, we need a mechanism for telling us which initial investments are most easily adapted, which investments can be revised, and how they should be revised if conditions change in one way or another.

This mechanism is contained in the notion of a strategy.\(^1\) A strategy is defined as a staged sequence of conditional decisions. Instead of specifying a single sequence of investments to reach a

---

\(^1\)See Massé [48] for an interesting historical summary of the notion of a strategy. We will define this term more explicitly in Chapter 2.0.
given end-state, the suggested procedure is to develop a series of conditional plans which express a series of "best" actions conditional on the history preceding that point. For example, if demand is high at the end of the first stage investment, we may decide to expand a two-lane highway to a four-lane; if demand is low, we may proceed to the second stage with no additional investment planned.

With such an approach, the first decision is the only one irrevocably committed. At any point in the implementation of the sequence, the plan can be revised or altered to meet changing or unexpected patterns of demand. Negatively stated, the objective now is not to search for the best sequence of investments and then implement the first stage of that sequence; rather, we are now looking for the optimal strategy (or series of conditional decisions) from among a number of strategies, and we then implement the initial stage of that strategy. The second stage decision will then be determined by the outcome of the first stage. If demand is high, we proceed with the high capacity alternative \( \Lambda_k \); if it is low, we construct a low capacity alternative \( \Lambda_m \). In terms of planning investments the primary concern now is not so much with future decisions as it is with the futurity or future impacts of the initial decision [49].

In terms of implementation policies, the strategy approach allows us to avoid committing ourselves to future actions or master plans until that time actually presents itself. Transport plans for

\[1\text{In principle, even the initial decision can be altered or scrapped for some cost—but normally the cost of doing so is prohibitive.}\]
1990 need now only be firmly committed five or ten years before implementation.

A Framework for Analysis

A suitable framework which recognizes both the sequential nature of investment decisions and the uncertainty of future demands is that of the general sequential decision model [50,51,52,53,54]. In its most general form, it can account for both interdependence between transportation and the socio-economic system as well as uncertainty and staged sequences of investment. At each stage of the analysis, it provides for the options of immediate investments, which include both staging and postponing of certain links, and the option of collecting more information.¹ We can represent the sequential decision model in extensive form as in Figure 1-4.

For each time period, t, there are a number of potential network investments (or actions), \( \Lambda_{mt} \), and for each network, there are a number of uncertain state variable levels (such as demand levels), \( \phi_{kt} \). The probability of a specific \( \phi_{kt} \) occurring, \( p_{kt}(\phi_{kt}) \), is assumed to be known. Any specific \( (\Lambda_{mt}, \phi_{kt}) \) pair is defined as an action-state variable couple. Each couple results in a set of consequences (e.g., network flows) by which the network may be evaluated for that particular point in time. In its most general form, the probabilities of the

¹See Johnson [55] for an exploratory research effort on information acquisition models for transport planning, and Jaramillo-Rego [56] for the use of a Bayesian approach for predicting the demand for public transportation.
Figure 1-4
The General Sequential Decision Model in Extensive Form
state variables $\phi_{kt}$ may be dependent on the value of $\phi_{k,t-1}$ at the previous stage or on a sequence of $\phi_{kt}$ at prior stages ($\phi_{k1}, \phi_{k2}, \ldots \phi_{k,t-2}, \phi_{k,t-1}$).

In principle, the logic of computation for deciding on initial investments is relatively straightforward, as we shall see in Chapter 3.0. However, in practice, the computational procedure for applying the model to transport planning is considerably more complex for a number of reasons. As stated by Manheim, et al [57]:

"First, there is generally a large number of combinations of actions and events. Second, the probabilities at different stages of the decision tree are different because information depends upon which actions were taken at earlier stages. Third, the utilities of a future period are different from the utilities at the initial stage. Fourth, and perhaps most significant, to evaluate the utility at any point in the decision tree may require running a complex simulation model, such as the urban transportation package (or other network equilibrium prediction models)."

Clearly, the use of this approach is then at best impractical, and perhaps even impossible, in a straightforward application, for thousands or even hundreds of points in the tree--or for any problem with a reasonable number of actions and events at each stage. Therefore, in order to apply the general sequential decision model to transport planning, it is necessary to develop special techniques adapted to the transport problem.  

---

1 See Manheim, et al, [58] and Hanson [59] for a preliminary discussion of these issues. See Hutchinson [60] and Khan [61] for a simple application of the statistical decision theory approach to transportation planning in which these issues are not addressed.
With that relatively brief introduction to the problem, let us now turn to the specific purpose and objectives of this study, followed by a description of the contents of the report.

1.2 Purpose and Assumptions

In order that investment decisions fully reflect a complete analysis and evaluation of alternatives, both present and future, predictive models of the transportation and economic system must be capable of predicting impacts and interactions over multiple time periods. In addition, they must be capable of analyzing the effects of uncertain estimates of specific model parameters on investment decisions. Most existing planning approaches have ignored uncertainty on the grounds that the system being modelled is complex enough as it exists, and represents a good approximation to the real world. In actual fact, a deterministic framework cannot produce a strategic investment pattern, in the sense that a good strategy would be responsive to changing information and changing conditions.

The principal objective of this study, therefore, has been to explore the problem of transportation investment planning when there are conditions of uncertainty over future possible outcomes. Essentially, it is a study of the where and when of transport investment expenditures which have characteristically long economic lives, and hence fairly significant uncertainty over future consequences. In a practical sense, the emphasis of this study is not with specific transport investment policies, however, but with the methodology of transport investment planning under conditions of future uncertainty.
Notably, our approach is similar to other problems involving investments over multiple time periods, such as the capacity expansion models of water resources [62] and the inventory models of the operations research literature [63,64]. Yet, it is unique in a number of respects, primarily because of transportation's overwhelming dependence and influence on the socio-economic system, and the complexity of the simulation procedure used to predict the flows and impacts of alternative networks.

Within this general problem area of investment planning under uncertainty, we have been concerned with the following four major areas of interest:

a. **The Effect of Uncertainty on Optimal Investment Patterns.** First, we have been primarily concerned with the issues of investments when faced with uncertainty over one or more demand parameters, i.e., the extent to which uncertainty may influence decisions. For example, tradeoffs in economies of scale versus savings from the partial staging of facility expansions are affected by whether or not future conditions are deterministic (or can be assumed to be so) or uncertain. After developing a model for capacity expansion under uncertainty, we explore the differences in investment patterns for deterministic versus stochastic conditions for the case of project design. These and other issues are analyzed and compared for differences in investment patterns.

b. **Solution Procedures for the Sequential Decision Model.** Second, the inclusion of the above factors and our decision to use a descriptive, nonanalytic simulation model for estimating returns from
alternative investment programs results in a sequential decision tree of investment-outcome combinations of extreme dimensions. Therefore, we must consider the problem of solution procedures. What are the alternative computational solution techniques and which are more likely to produce good alternatives? Clearly, the resources required for enumeration of thousands, or even hundreds of points in the tree will require careful and selective pruning of alternative branches. Sections 4.2 and 4.4 discuss these issues in more depth. We have, then, also focused on the basic procedure of efficiently searching for time-staging strategies when using a network equilibrium model and a stochastic demand structure.

c. **Adaptive Learning about Probabilities.** Third, we have also extended the general sequential decision model in the extensive form to allow for the case when the information about the underlying probability distributions is to some degree uncertain. This extension is an adaptive mechanism (using a Bayesian updating procedure) for the sequential decision model which allows the planning process to learn from previous actions about the true nature of the underlying uncertainty.

d. **Project Versus Network Issues.** And finally, the ultimate objective of this research is to be able to analyze efficiently time-staged strategies of network investments in the face of uncertainty. In order to deal with the combinational problem associated with the network case and yet keep our computational efforts within reason, our research strategy has been to focus primarily on the project design
level, but with simple network effects included. Thus, the results from exploring the first three objectives pertain to project analysis. With the understanding developed from these, we hopefully will be able to develop insights into how to handle the large network problem efficiently.

To accomplish these objectives, we have implemented a series of computer programs and conducted some initial experiments to explore the concept of the sequential decision model as applied to transport investment planning. The programs are designed to both simulate the supply-demand network flow patterns at a single point in time, as well as produce shifts in demand over the longer run for conditions of an uncertain or stochastic demand formulation. The implementation examples are hypothetical since it would have been at best impractical, and perhaps even impossible, to collect all the relevant data, calibrate the models, etc., in the time available for this study. We have also made some extremely simplifying assumptions, because of the scope of the investment problem, about demands and supplies and we have limited our examples to a single technology, i.e., there are no modal dependencies. Including a number of modes is a natural first extension, but since it should have no appreciable effect on our proposed procedure, we have chosen to ignore it in order to achieve additional savings in computational time. On the other hand, we have attempted to be as realistic as possible in our modelling of network behavior by incorporating a supply-dependent variable demand structure.

In the framework of larger studies (Manheim [65], the Harvard-Brookings models [66], de Neufville and Stafford [67], and so on),
we recognize that investment planning models are only a small part of the overall process of the analysis of investment alternatives. The problem of investment in transport facilities involves a number of issues that cannot be captured fully by any model of the transport and economic system. For example, multiple and conflicting objectives, externalities, and political influences are difficult, if not in fact impossible, to model in any realistic fashion. On the other hand, a purposeful, directed effort at systematically analyzing investment alternatives using computers and the latest systems analysis techniques can lead to powerful insights into some of the more subtle issues of investment. It is in this spirit that we present the material of this thesis as a contribution to the analysis techniques of transportation planning.

In summary, the focus of this research is directed toward exploring the problem of the time-staging of transport investment in the face of future uncertainties. One result of this exploration has been the development of a set of computer programs that model the transport and economic systems as a time-staged investment decision problem. The model can, theoretically, handle all of the dependencies described in Section 1.1, in addition to explicitly recognizing the underlying uncertainty surrounding investment decisions and the true sequential nature of transport investments. Using this model, we have been able to conduct some experiments of an exploratory nature into:

(1) the effects of uncertainty on investment patterns for the project design case,
(2) the feasibility of applying the general sequential decision model in extensive form to the transport investment problem, and

(3) extending the basic capacity expansion under certainty framework to include an adaptive learning procedure.

1.3 Structure of the Study

In the preceding section, we described the basic purpose and assumptions of the study. In order to accomplish these objectives, we must first review the existing planning methodology and the supply-demand paradigm of economic theory. We then present the framework of the sequential decision model and some possible solution techniques for it. After presenting the proposed heuristic procedure, we then develop the experiments and analyze the results which are designed to accomplish the four major objectives of this study.

More specifically, Chapter 2.0 provides the overall statement of the problem by presenting the basic problems of capacity expansion or investment time-staging for the two cases of project design and program selection. It defines the transportation investment problem in terms of the basic supply-demand paradigm of economic theory and relates this to the current set of procedures and models used for investment planning. With this background on current approaches, we then turn to the incorporation of uncertainty and the general sequential decision model as a solution framework. Finally, we present a survey of some current transport time-staging problems and the solution techniques used in these studies.

Chapter 3.0 describes a number of promising search techniques for this sequential decision framework, such as dynamic programming,
branch and bound and integer programming, monte carlo simulation, etc., and their advantages and disadvantages for the transport investment problems we have defined. In this chapter, we also present the procedure we propose for solving the transport time-staging problem, along with its limitations and drawbacks and suggest a means to get around these constraints.

In Chapter 4.0, we present the proposed set of heuristic procedures for the transport time-staging problem under uncertainty using what is now commonly called decision analysis as the basic model structure. This section defines, then describes in detail the various pruning rules and terminal evaluation functions needed to make the procedure feasible.

Chapter 5.0 then presents our computational experience with the proposed procedure using the set of computer programs (DECISN) constructed to test a number of hypotheses. The experiments, conducted for the project design case or simple single-link investment problem, explore two major but basically different kinds of problems. The first explores the more theoretical issues of staging and timing of transport investments. The examples are simple enough that computational difficulties we are likely to encounter with larger problems can be ignored. In these experiments, we explore the tradeoffs between opportunity losses from excess capacity and scale economy losses from staging under assumptions of deterministic, and static and dynamic uncertainty conditions. We also present the results of experiments performed on the relationship between adaptability, staging, and
uncertainty. The second major set of experiments in this chapter is directed toward comparing the proposed search procedures, i.e., decision analysis with the approximation procedures of pruning rules and terminal evaluation functions, for a number of conditions.

Finally, we extend the basic capacity expansion model in Chapter 6.0 and develop an adaptive learning model based on a Bayesian procedure which allows us not only the flexibility of strategic investments, but to learn about our underlying demand model uncertainty. We first outline the traditional Bayesian decision theory framework and the difference between the traditional model and our multistage investment model. (Appendix D presents an extension of this Bayesian technique which also allows us to learn about the underlying conditional distributions.) We then present the results of a series of runs using the adaptive version of the model and compare these results with the results of the previous chapter. We conclude the chapter with a brief summary and the conclusions reached from these experiments.

In the concluding chapter, we summarize the results of the research effort, the conclusions reached from the limited set of experiments conducted, and finally, the implications on other research areas, and future extensions of this effort.
Chapter 2

THE TRANSPORT INVESTMENT PROBLEM
Chapter 2

THE TRANSPORT INVESTMENT PROBLEM

Chapter 1.0 has briefly outlined our area of interest in transportation, some of the problems involved in long-range planning, and the general purpose of this study. The purpose of the present chapter is to review the existing approaches to transport planning and to present a more detailed and explicit statement of the problem.

Section 2.1 first describes the general methodology of transport investment planning as it has evolved over the last decade or so and attempts to place this study in perspective with a larger, ongoing research effort. With this background, Section 2.2 then presents a specific statement of the problem in terms of the issues outlined in Chapter 1.0. Section 2.2.1 first defines the basic problem of timing, staging and economies of scale for the simple case of project design. In Section 2.2.2, we extend this to include network and supply-demand dependencies for the case of program selection. In Section 2.2.3, we discuss the consequences of introducing uncertainty into the staging problem and in Section 2.3, the general sequential decision model is presented as a proposed framework for the transport investment problem under uncertainty. Finally, Section 2.4 presents a survey of the literature and Section 2.5 summarizes the chapter and discusses the conclusions reached from our survey.
2.1 The Nature of Transport Investment Planning

The problem of specifying a transportation investment policy is obviously considerably more complex than we have indicated in the previous introductory chapter. To adequately specify an investment policy involves much more than merely accounting for the dependencies described earlier. This section is intended to review the nature of transport investment planning and to describe the general equilibrium framework which is the basis for the models we will use in this study.

2.1.1 The Objectives of Transport Investment Policies

A first consideration in deciding what capacity, where and when it is to be implemented is the question of objectives—how do alternative transport investments, mixes of technologies, scales and timings meet the objectives set down for that policy.

Even a very cursory review of the major urban transportation studies, the economic literature and developing country transportation studies will reveal that very rarely have objectives been spelled out in enough detail so this question can be answered satisfactorily. For the most part, existing transport studies have focused primarily on those ends which can be classed as micro-economic in nature and can be defined in some detail. Many of the early pre-urban transport studies were undertaken by engineers or engineering-oriented agencies who were concerned largely with the micro-economic and engineering details of design speeds and operating, maintenance and fixed capital costs. Even existing transport studies have in many instances
limited objectives to the very micro and detailed variables of transportation hardware.

Transportation is very rarely desired for its own sake, however. Its impact on the environment is usually much more pronounced than what an improved operating speed or decreased operating cost would indicate. It is more generally an instrument used to promote other ends—ends which are very broad in scope and difficult to define operationally. For example, transportation determines to a large degree the factor of mobility—it allows both men and material resources to shift to areas where they can be employed most effectively. It also enables both goods and passengers to be transferred between production and consumption centers; it changes relative factor costs through increased speeds, lower operating costs and general improvements in safety factors [1], ..., and so on. Lansing [2] has defined this relationship between transport and economic policy objectives even more broadly:

"Among these goals (of economic policy) are economic efficiency, economic growth, a high level of employment and freedom from pronounced cyclical fluctuations, and a degree of equity in the distribution of the products of economic activity which avoids the juxtaposition of extreme poverty and extreme wealth. The transportation industry is directly involved in the attainment of these objectives."

Thus, while the traditional engineering economic and highway investment literature suggest that the problem of making investment decisions (including specifying the criteria for choice) is relatively
straightforward, a number of recent studies concerning the relationship of transportation to other sectors of the economy would suggest otherwise.

To further complicate the problem, however, there are usually a number of other objectives besides the economic ones which are just as general and just as difficult to implement operationally. Fromm [3] states that transport "serves to increase national defense capabilities, social cohesion, and political stability." In the past, often all of these objectives (economic and non-economic) have been conflicting and extremely difficult to reconcile. In these cases, generally one objective has been selected as the primary objective with little if any consideration given to the complementary and competing effects on other objectives. The interstate highway system is a prime example of a solely politically motivated investment decision.¹

Fortunately, agencies concerned with investments of a transport nature are beginning to recognize the complexity of transportation objectives. Competing and multilevel economic and non-economic ends

¹Justified on the basis of national defense policy, it has been shown, however, that the system, even though not necessarily optimal, is much preferred to no improvement from an economic efficiency point of view [4]. The author also suggests that with some economic analysis, it might have been possible to design a system even more beneficial which could serve both urban and rural interests and still provide an integrated system for defense purposes. In other words, it is possible to consider both economic and political objectives.
for which transport investment can serve have received considerable recent emphasis in the literature. An appraisal of urban transportation planning was completed recently by an international organization, the Organization for Economic Cooperation and Development in Paris, of which some 21 countries, including the United States, are members. A panel of experts prepared a report on "The Urban Transportation Planning Process: In Search of Improved Strategy" (OECD, 1969) [5]. The conclusions of this panel summarizes the change in emphasis that has taken place concerning investment planning objectives and strongly emphasizes the need for a changed planning methodology.

"Depending on the skill with which we exploit its potential, transportation can be either an instrument of desirable social change or a disruptive force against human development. It can both enhance and damage the quality of the environment. It can act either as a stimulus or as a brake on urban growth and development. Thus the ability to make enlightened transportation decisions may to a large extent determine government's success in achieving wider policy objectives.

Two premises served as foundation for the meeting. The first premise was that a new conceptual approach to urban transportation planning is emerging - one which gives increased emphasis to human values and to the social and economic goals of urban development. In this approach economic and engineering efficiency, 'demand' for transportation, and profitability no longer serve as the only guiding principles for investment decisions. These conventional criteria are weighed against the social, economic, environmental and aesthetic needs of urban residents: personal mobility, accessibility to urban opportunities, comfort and convenience, clean air, open spaces, pleasing surroundings, and preservation of neighbourhoods and of urban diversity. Underlying this shifting emphasis is the growing conviction that transportation is not an end in itself but a tool for bettering the total condition of
urban life; that its objective is not just to move people but to enhance the quality of the cities and to improve the social well-being of their residents; and that planning concerned only with the effects on transportation itself has too often resulted in transportation systems that have failed to contribute effectively to these objectives.

A concomitant premise was that there is a need for a methodology which is more sensitive to the important issues facing urban society and more effective in helping to reach socially responsive decisions. In particular, more sophisticated tools of analysis are required to perceive individual and community preferences and formulate goals and programme objectives in the light of evolving technology and changing habits and values; to search for and generate alternative approaches to meet given objectives; to predict, evaluate and rank the impacts of alternative proposals; and to give adequate recognition to the element of uncertainty in the design of decisions."

2.1.2 The General Planning Methodology of Transport Investment Planning

In response to a changing attitude towards the purpose of transport investments and the nature of the planning process, there have been a number of studies that have at least implicitly recognized this complexity and conflicting nature of transportation objectives and the need for a changed planning methodology [6,7].

A very general procedural framework which recognizes this changing emphasis in planning methodology has in fact been developed and partially implemented [8,9]. In its most general form, it recognizes the fact that not only are demands and technology changing, but our goals which encompass economic and non-economic objectives, are also changing. Although the details of this model are beyond the
scope of our present research, we will present the portion of this work most relevant to our current purposes while recognizing it is part of a larger more flexible planning procedure. This portion of the larger framework is the *general equilibrium framework* of transport investment planning.

**The General Equilibrium Framework**

The underlying premise of this framework is that the transport systems analysis problem can be modelled in terms of two important broad classes of variables:

(1) transport options

(2) activity system options

Transport options are those items that can be controlled directly by transportation planners. They are the decision variables which range from such broad items as alternative technologies to specific hardware items such as networks, links, and vehicle types, to softer items such as operating policies. Activity system variables are those social, political and economic variables which determine the demand for these transportation options. They include direct variables such as the spatial patterns of population and economic activity, and the more indirect variables of mortgage policies and housing subsidies and the like, all of which can influence in one way or another the demand for transport services. In some cases these options can also be controlled

---

1The discussion of this section broadly summarizes the major points of two documents of a larger related study effort [10,11].
by the decision-maker. In most instances, they are taken as exogenously specified; non-manipulable in the direct sense, but still considered as a major factor in the analysis.

Given these two sets of variables, the core problem of transportation systems analysis in the general equilibrium framework is defined as the prediction of flows on the transportation network i.e. the interaction of any specific investment option, T, with the activity system, A, results in a unique set of transportation network flows. The general framework used to predict these flows is the equilibrium of supply and demand [12,13,14]. Transportation options, T, can be specified and differentiated more precisely by a set of supply functions, S; activity system variables by a set of demand functions, D. The equilibrium that occurs between these two sets of variables produces a set of network flows, F, which depends on the specific set of transport options in place and on the unique characteristics of the activity system at that time.

This general description can be represented more concisely by the following vector notation:

\[
T = \text{specification of the transportation system in terms of the full set of available options - technologies, networks, vehicles and operating policies.}
\]

\[
A = \text{specification of the activity system, in terms of both exogenous characteristics such as national population and economic trends and controllable options such as land use controls.}
\]
\[ F = \text{pattern of flows of passengers and/or freight in the transportation system.} \]

\[ L = \text{service characteristics of a particular flow or set of flows - travel times, fares, comfort, etc.} \]

\[ V = \text{volume of flows on the transportation network.} \]

These variables are then used to specify (a) supply functions, \( S \), which give the level of service in terms of trip time, trip cost, comfort, and so on, as a function of the transportation options and the volume of flows, and (b) demand functions, \( D \), which give the volume of flows \( V \) as a function of the activity system options and the level of service:

\[ L = S(T,V) \]

\[ V = D(A,L) \]

The equilibrium point, or intersection of these two functions, results in a set of flows, \( F \), characterized by specific values of flows and the levels of service, \( V \) and \( L \):

\[ F = (V,L) \]

A simple two-dimensional graph shows this procedure schematically (Figure 2-1). Both \( V \) and \( L \) are assumed to be one-dimensional in this diagram - more generally, they are multidimensional vectors.

The use of this formulation for predicting system changes and evaluating improvements is described simply by Figure 2-2. The current flow pattern is defined by the equilibrium of \( S' \) and \( D' \), or point \( E' \). Considering a system improvement \( S'' \) results in a new equilibrium \( E'' \), and a new set of flow volumes and a new level of service.
T \iff L = S(V,T) \iff F = (V_0, L_0)
A \iff V = D(L,A)

Figure 2-1

Simple Equilibrium
Figure 2-2
Changing Equilibrium Due to System Improvement
Given criteria for choosing among alternative improvements, it is a simple task in principle, to evaluate and compare alternative improvements and choose the alternative which gives the most improvement for the amount invested. This then is defined as the static problem of transport systems analysis.¹

The Use of the Supply-Demand Paradigm for Changes over Time

Because transportation investments are extremely long-lived and can heavily influence the activity system changes in terms of spatial patterns of demand and economic activity location, this prediction must also account for the time dimension as well as we indicated in Chapter 1.0. This is represented in principle in the supply-demand diagram by a shift in demand to a new position D" (Figure 2-3(a)). In practice, this change represents a shift in both the magnitude and spatial distribution of demand in terms of population and economic activity from one period to the next.

In two dimensions, we can represent the shift in demand by a shifted demand function, D", which results in a new flow pattern, F", if no new supply (i.e. no new links, networks, etc.) has been implemented between period t and period t+1 (Figure 2-3(a)). This

¹The prediction of flows in a network is based in principle on the theory of equilibrium between supply and demand. In practice, because of the difficulty and expense, this computation is often approximated by assuming inelastic demand, constant supply functions, etc. For a discussion of the errors involved in this approximation and the actual mechanism of calculating equilibrium in a transport network, see Ruiter [15].
Figure 2-3
Extending the Supply-Demand Paradigm to Multiple Time-Periods
58
new demand curve represents the normal growth in population which produces a new equilibrium flow volume.

If investment does occur, there are two possible changes that can occur in the pattern of equilibrium flows and level-of-service. The first is represented in Figure 2-3(b) by the intersection of the demand function $D''$ with the new supply $S''$ yielding an equilibrium flow volume $V''$. This represents the case of long-run supply-demand independence i.e. there are no shifts in demand due to increased capacity and accessibility, only normal population growth shifts in demand.

The second case is shown by the intersection of demand curve $D'''$ with $S''$ yielding flow volume $V'''$. This case represents a demand shift due to transport improvements, i.e. there are shifts in population because of improved accessibility, and a change in volume results because demand itself is a function of the capacity provided.\(^1\)

The use of this basic supply-demand paradigm as a prediction mechanism for transport flows and impacts, with demand shifts represented by a demand shift function, is shown schematically in Figure 2-4. It involves specifying options of both transportation and the

\(^1\)We must be careful to note that such positive shifts in demand to one region are accompanied by negative shifts in other regions. The net effect is difficult to determine, a priori. The important thing to note is that such shifts must be explicitly accounted for.
BASIC PREDICTION MODELS

Figure 2-4

Basic Prediction Models

60
activity system which are then "modelled" in terms of supply and demand functions. These are used in a complex iterative simulation procedure to predict the impacts for a wide number of groups concerned with transport consequences. These impacts along with the equilibrium flows and changes in accessibility are then used to determine changes in demand for subsequent periods.

The Equilibrium Framework in Practice

In principle, the computational sequence used for the general equilibrium framework is simple and straightforward. However, in practice, the prediction of equilibrium flows in a network is difficult and costly to achieve at best. In fact, although all of the major existing transportation studies can be described in terms of this framework, it is only in the most recent studies that the full interaction of supply and demand over multiple time periods has been attempted [16,17,18]. Even these more recent approaches involve a fairly complex, iterative and ill-defined sequence of computations with their own biases toward the equilibrium computation.

2.1.3 Scope of this Study

The intent of the previous sections of this chapter has been to describe briefly the overall framework for transport investment planning; to recognize the complexity of both the objectives of transport investment policy and the methodology of transport investment planning designed to achieve those objectives; and to point out that
in fact, there is no one methodology that can define the overall process or no one set of techniques that will produce a best set of plans. Rather, the process involves a complex set of models, techniques, and procedures to be used in a very flexible and iterative manner to gain insights into broad questions of investment policy and more detailed questions of specific investment plans. It must also recognize the social, economic and political constraints that exist in arriving at both policies and specific plans.

In order to focus this study, however, our primary emphasis has been restricted to the development of a methodology of predicting, evaluating, and choosing time-staged investment strategies in the face of future uncertainties for a fairly well-defined problem which is a subset of the larger transport problem we have just described. Therefore, although we recognize the complexity of the problem, we have restricted our initial study effort to questions of sequential decisions in an economic efficiency sense. Our general model and search procedures, however, reflect the fact that we consider our procedure to be part of this larger process and that, hopefully, we will be able to incorporate them in the prototype system based on this framework in the very near future. Consequently, we have not concerned ourselves with specific optimization techniques such as mathematical programming nor have we concerned ourselves with the accuracy and details of existing prediction models. Our primary emphasis, in fact, has been on a very general heuristic approach to developing time-staging strategies using existing transportation prediction models. Thus, while we use an expected value maximizing principle in choosing among
strategies, we are more concerned with the philosophy of sequential
decisions, the strategy approach, and the flexibility and adaptability
of investment plans.

An initial operational model [19] based on the general equili-
brium framework has undergone preliminary testing and presently is being
extended. A number of preliminary experiments using this model were per-
formed by the author [20] to explore the implications of time-staged in-
vestment strategies. Unfortunately, developing time-staged investment
strategies proved to be too expensive and too cumbersome for our explora-
tory work with such an elaborate set of prediction models. Therefore,
recognizing this framework, but using a simpler set of prediction models,
we developed the current set of models (DECISN) in order to test the
specific emphasis of this thesis—the effects of uncertainty and sequen-
tial decisions in a planning framework.

We now turn to a statement of the problem as defined for this
thesis, and a survey of the work that has been done in this general
area of investment planning and prediction of time-staged investment
strategies.

2.2 Statement of the Problem

Recognizing the framework of the previous section and the
problem characteristics of Chapter 1.0, we can summarily state that
the basic transportation investment problem is to determine an invest-
ment plan that maximizes overall community or national welfare objec-
tives while explicitly accounting for the following characteristics:

(1) supply-demand interdependencies
(2) network connectedness

63
(3) the multistage nature of investments

(4) dependencies due to longer-run supply-demand shifts
(timing dependencies)

In practice, this approach has not been actually implemented in any existing study in its full form. Many approximations have been made in both the models employed and the actual methodology of using them.

But more importantly, the general planning framework, both as we have presented here and as it is being used in practice, is not equipped to deal with either:

(1) the fact that there are major uncertainties over activity system variables (demand), their magnitude and growth, and over transportation technology itself (supply) in terms of operating costs and performance of the system under a variety of loads; or

(2) the sequential nature of investment; investments are implemented as staged sequential increments to an existing system and, although the general framework can be used to plan multistage investment plans, it is severely constrained in the number of alternatives it can realistically consider.

A number of planning approaches have been developed, however, which have focussed on investments over time or the multistage nature of investments for conditions of a deterministic future. Before presenting the general sequential decision model as a framework for handling uncertainty and the sequential nature of investments in Section 2.3, we will first present the problem of transport time-staging in terms of these existing methods, emphasizing the differences that exist between the general transportation investment problem and these existing approaches.

2.2.1 The Issues of Project Design and Analysis

Historically, the problem of project design and analysis can be
described as perhaps the most traditional method of planning, in the transport, as well as other fields. It generally requires the least amount of computational effort of all planning approaches but only because it chooses to decide on expansion capacity at the very lowest level of a project, ignoring dependencies with any other part of the system.

For our purposes, a project is defined as a specific investment alternative which is usually relatively small in scale and assumed to be independent (whether it is or not) from other projects. It may be a specific plant location, an airport, or a proposed link in a highway system. The problem then is to design, predict the consequences, and choose between several project alternatives which are assumed to be mutually exclusive. This is commonly referred to in the literature as the problem of project design [21].

The assumption of independence from other projects implies, for the highway link example, that (1) there is no dependence with other links in the network, or that if there are, this dependency can be explicitly accounted for, or that it is negligible and can be assumed away, and (2) there is no financial dependency between projects, i.e., projects are not competing for the same funds or budgets.

The second major factor of importance is that projects generally have two associated costs, a fixed and variable portion. The fixed portion represents the capital cost and is independent of the amount of use or output of the link. The variable cost depends on the use of the facility. The decision to select a specific design is then a decision between alternatives with different fixed and variable costs (see Figure 2-5) or of different scales.
Figure 2-5

Multiple Scales of Transport Investment
The third major factor is related to the second; although projects are assumed to be mutually exclusive (spatially), a project generally has consequences that extend into the future, i.e., transportation projects are durable goods with an associated economic life. Therefore, when deciding on a project design, the decision of what scale facility to provide depends on the costs attributable to the different scales available for a number of different periods; in other words, because a demand is usually growing, for both low (or initial) and high (or subsequent) demands. Large-scale facilities have excess capacity at low volumes. Therefore in terms of the opportunity losses associated with scarce resource problems, they will be charged with large losses due to unused capacity at low volumes. Conversely, smaller scale facilities will have smaller losses at low volumes but relatively larger losses from being under capacity at high volumes.

Multiple Capacities and Economies of Scale

Two additional factors enter to complicate the problem. First, projects need not be either-or questions. That is, although concerned with present and future demands, the decision is not usually restricted to one of a number of mutually exclusive scales. We can, in the most general capacity expansion case, add capacity in small increments, matching capacity to demand, to minimize the losses from excess capacity. Therefore, when deciding on a project design, the initial decision depends in part on the initial demand and the cost of a specific facility and, in part, on later demands and the costs of additional capacity. Thus an initial selection may depend on the availability and cost of future
expansion.  

Assuming demand to be exogenously predicted over the period of interest and capacity and time to be continuous variables, the problem reduces to selecting the times $t_j$ when particular levels of capacity $\Psi_j$ are to be implemented (see Figure 2-6).

The second factor is that, generally, a larger project costs proportionately less to implement than two smaller projects of the same total capacity and has a smaller per unit variable cost than either of the two smaller alternatives. This is defined as the effect of scale economies. We can therefore choose a large project now which will have higher operating costs at the lower volumes and overall lower fixed costs, or the two smaller project implemented in sequence, which will have low operating costs initially, and an overall total higher fixed cost. The decision will depend on the actual values of the costs, the discount factor used to discount future costs, and the expected growth in demands.

**Application to the Transport Problem**

A number of simple models have been used in one form or another for a great many capacity expansion problems. Chenery [24] develops a simple capacity expansion model for exploring the effects of scale

---

1 The "cost" of future expansions, in turn depends on how society chooses to value future vs. present consumption, i.e., what is the discount rate used to account for the fact that a dollar today is worth more than a dollar in the future. For a complete discussion of discount rates representing society's time preference for consumption, the opportunity cost of capital discount factor, and combined synthetic rates, see Marglin [22] or Wohl [23].
Figure 2-6
Two Typical Investment Sequences for a Single Demand History
economies for pipeline investment. For linear growth in demands, he shows that the optimal scale and time of expansion is a constant. Manne [25] has simplified Chenery's formulation and extended the results for a stochastic demand structure and also for non-linear demand growth [26]; these formulations are similar to much of the literature in inventory control [27]; Muhich [28] and Russel [29] have developed versions of this kind of model for water treatment plant capacity expansion. Marglin [30], Tarplay and Drake [31], and Thygeson [32] also present similar formulations which are concerned with the timing dimension and possible postponement of investment projects.

The principal characteristics of all these formulations are the following:

(1) demand is assumed exogenously predicted and independent of the alternative selected.

(2) project capacity is merely additive; to meet increasing demands, we need only add another scale to the project and both now carry the total volume.

(3) costs are very simple in structure; generally a fixed and a variable portion, and the life of the facility is essentially infinite.

(4) the problem is one of minimizing total costs only—of balancing off the costs of being under and over capacity.

(5) attention is usually focussed on the analytical structure of the problem and a closed-form solution technique is used to produce an optimal investment sequence.
The transportation investment problem, though similar in principle to the capacity expansion problem, is considerably more complicated because of the following unique characteristics:

(a) Demand

The demand for travel, as a short-run derived demand function, has been the subject of considerable attention over the last few years. Unlike most of the capacity expansion formulations, transportation demand cannot be taken as a fixed exogenous quantity. Generally, it is not independent of the supply (or capacity) offered, nor independent of the time of day. Normally it is a downward sloping function of the price or cost variable and is unique for different times of the day (see Figure 2-7).

The formulation presented here is similar to the demand models to be used in Chapter 5.0 and is based on the demand model known in the literature as the SARC-KRAFT model [33]. Essentially, demand is stratified by mode, trip purpose and time of day and is taken to be a function of (1) the socio-economic characteristics of the region, and (2) the service levels or "price" of the supply between two zones. The model generally takes the following form¹ (excluding time of day):

\[ v_{ij}^m = \alpha_0 + \alpha_1 m(Y_i) + \alpha_2 m(E_j) + \alpha_3 m(p_{ij}) + \beta m(t_{ij}) + \gamma m(f_{ij}) \]

\[ \ldots (f_{ij}^m)^{\theta m} (p_{ij}^n)^{\beta mn} (t_{ij}^n)^{\delta mn} (f_{ij}^n)^{\theta mn} \]

¹A more general formulation would also be stratified by time of day and include cross-elasticity terms for service variables for different times of the day. See Wobbi [34].

71
Figure 2-7
Intra-Temporal Demand Functions
where

$V_{ij}^m = \text{the demand for travel between } i \text{ and } j \text{ by mode } m$

$\alpha_m = \text{parameter for travel by mode } m$

$P_i, P_j = \text{product of the populations at zones } i \text{ and } j$

$Y_i = \text{average per capita income at } i$

$E_j = \text{employment opportunities at } j$

$p_{ij}^m = \text{fare or money price by mode } m$

$t_{ij}^m = \text{travel time by mode } m$

$f_{ij}^m = \text{frequency of service by mode } m$

$p_{ij}^n = \text{fare or money price by mode } n$

$t_{ij}^n = \text{travel time by mode } n$

$f_{ij}^n = \text{frequency of service by mode } n$

$\alpha_{1m} = \text{percent change in mode } m \text{ travel resulting from a 1\% change in population product}$

$\alpha_{2m} = \text{percent change in mode } m \text{ travel resulting from a 1\% change in income}$

$\beta_{mm}, \delta_{mm}, \theta_{mm} = \text{percent change in mode } m \text{ travel resulting for a 1\% change in one of the service variables of mode } m$

$\beta_{mn}, \delta_{mn}, \theta_{mn} = \text{percent change in mode } m \text{ travel resulting for a 1\% change in one of the service variables of mode } n$

The parameters $\alpha, \beta, \delta, \theta, \text{ etc.}$, represent elasticities and cross-elasticities or the responsiveness of demand to changes in the independent variables. More specifically, elasticity is defined as the percentage change in demand expected from a 1\% change in "price," where price may be any one of the service variables. Cross-elasticities are defined similarly but reflect the change in demand for one good, highways for example, when a 1\% change in price is experienced for another good, e.g., mass transit.
More generally, the form can be presented as

\[ V_{ij}^m = f(S, L) \]

where

- **S** = the vector of socio-economic characteristics of the zones,
- **L** = the vector of "price" or service level characteristics of all modes

with a separate equation being developed for each trip purpose, such as work, shopping, business, school, etc. Each equation will then have different independent variables and different elasticities.

(b) **Cost Structure of Technology**

Similarly, costs are not so simply defined for transportation. There are at least three major differences in the nature of costs between the simple capacity expansion models and a transport planning problem:

1. **the "costs" of transportation include much more than**
   - two simple fixed and variable portions that are to be **jointly minimized**.

The form of the variable cost structure or the "supply" function the user perceives is similar to the traditional average variable cost function of micro-economic theory. Figure 2-8 shows these functions, which are similar to those found in Wohl and Martin [35], for three basically different kinds of highway facilities. These supply functions represent the price of travel a user will experience as the total volume varies on the facility. As such, they include all explicit and implicit costs the user will encounter in making this trip, such as user tax payments, perceived vehicle operating costs, parking fee payments and travel
Figure 2-8

Technology Supply Functions
time. It also represents an implied value scheme or weighting function and imputed dollar value of non-dollar items. In other words, we are collapsing the level-of-service vector of the previous section to the single dimension of cost by placing dollar values on wait time, travel time, comfort, and so on.¹

In addition, the real costs themselves are comprised of a number of elements: fixed costs are comprised of fixed capital cost, right-of-way costs, and interchange costs (all with different economic lives).

(2) transport requires a significantly different treatment of multiple scales.

Multiple scales in the transport capacity expansion case can be achieved partially by partial construction of a particular facility; by building only two lanes of a four-lane facility. This is a much more subtle approach than the one taken by the existing capacity expansion models, however, since we can provide two lanes now and add two lanes later, build two lanes now with provision for right-of-way for another two, build two now with provision for two more plus add interchanges now, and so on.

(3) it also requires a more explicit recognition of the costs of expansion.

Most importantly, however, the costs of expansion are not as simple as defined by the capacity expansion models or as defined by the

¹For a discussion of this approach used in disaggregate modal split models, see Lisco [36], Lambe [37], Quarmby [38], Lave [39], or Wynn [40].
general long-run cost curves (the envelope of the short-run curves).

Walters has succinctly pointed this out [41]:

"The irreversibility of road investment enhances these effects. Once we have built a road it is sunk in a particular location for a long time. Furthermore, and this is where it differs from the majority of investment models, the irreversibility occurs in increasing the capacity of the roads. (In my view this irreversibility is typical of virtually all economic processes. The path of adjustment affects the final desired outcome; there is no unique long-run equilibrium—there are as many long-run equilibria as there are paths to it. This suggests that the general analysis of the long-run needs considerable revision—but, of course, this task is not attempted here...) For example, if the road authority had already built a two-lane highway, the cost of widening it into a four-lane highway will be different from the cost of a four-lane road if we were starting from scratch. There are as many costs of a four-lane highway as there are paths to it. Indeed, if for some reason a three-lane highway had been built, it might not be worth extending to a four-lane road; only if we are starting with a two-lane road may it be worth building the high-capacity four-lane motorway.

In one sense, the irreversibility argument is nothing more that the old adage 'bygones are forever bygones'. But applied in practice, it implies that each expansion path is defined in directional terms. There is no such thing as the minimum long-run cost curve. We have illustrated the example of the two to four-lanes and the three to four-lanes options in Figure [2.9]. The broken line ... indicates the two to four expansion path, and the continuous line ... shows the three to four expansion path. There are two fixed costs for the four-lane highway, $F_{24}$ and $F_{34}$, according to whether one begins with a two- or three-lane highway; but we have supposed the variable costs are the same."

In a real-world problem, the subtleties of partial staging makes this process even more complicated.

(c) **Equilibrium at a Stage**

In the case of most transport investments, and particularly highway investments, we also cannot focus solely on total costs of each invest-
Figure 2-9
Capacity Expansion Cost Paths
[After Walters]

78
ment or sequence of investments because of the dependence between supply and demand. There are costs (the link supply function) which are important in determining whether or not people will make a trip (which are used to determine equilibrium), and there are other costs (capital, vehicle ownership, etc.) which are used to evaluate whether that investment is a good one (but are not necessarily important to determine equilibrium).

The procedure of lumping both fixed and variable costs together ignores the fact there are different costs perceived by different interest groups—in our case, the users of the facility and the providers of the same facility. For example, the equilibrium flow volume in the highway case is determined by equating supply and demand using the cost variables as perceived by the user of the facility, not the total costs. Thus, the supply function only includes those costs that the user sees in making the trip, the average variable user costs of Figure 2-8. All other costs associated with the facility are ignored for the moment. This equilibrium is shown in Figure 2-10.

Once equilibrium has been determined for all stages, yielding the volume expected to use the facility and the cost they will experience, we may now turn to the inclusion of other fixed costs and the evaluation of different alternative sequences.

(d) The Measure of Benefits

There are several dimensions to the value of any investment which includes both benefits and costs attributable to any facility. The cost side we have discussed. On the benefit side, a transport improvement can generate benefits ranging from such simple and measurable items as user
Figure 2-10
Equilibrium with the Average Variable Cost Function
savings in operating costs and travel time to more indirect items such as intergovernmental transfers and ill-defined aesthetic values. A list of both potential costs and benefits attributable to transport, taken from Wohl and Martin [42], is contained in Table 2-1.

The primary difference between existing capacity expansion models and transport investment is that the former usually fail to measure the benefit side at all. Our measure of benefit in this study will concentrate mainly on user travel benefits in the form of willingness-to-pay.

Willingness-to-pay is defined as the price that consumers (travellers) are willing to pay for the service in question. For downward sloping demand curves, it often is more than what they have to pay for all travellers except the traveller at the margin; i.e., in Figure 2-11, all travellers up to the equilibrium volume, \( v_k \), are actually willing to pay more (according to the demand curve) than the equilibrium price \( p_k \). This is equivalent to a measure of gross benefits.

Consumer surplus is a more familiar term which is equivalent to the net benefit of any improvement. It is defined simply as "the difference between the maximum amount consumers are willing to pay for a specified quantity of a good rather than go without it and the value of the given quantity at the good's competitive marked price" [43]. Therefore, it is represented in Figure 2-11 by the cross-hatched area under the demand curve but above the equilibrium price \( p_k \).

(e) Aggregate Networks

The most significant difference between capacity expansion models and a transport time-staging model, however, is in the definition of ex-
Table 2-1
The Potential Benefits and Costs of a Transportation Investment
[AAfter Wohl and Martin]

Transportation System Costs and Benefits

A. Potential Costs

1. Facility constructions and land acquisition
2. Dislocation, social costs (externalities)
3. Facility operation, maintenance, and administration costs
4. User travel costs
   (a) Vehicle ownership (excluding fees and taxes levied to recover facility costs)
   (b) Vehicle operation and maintenance (excluding fees and taxes levied to recover facility costs)
   (c) Time costs
   (d) Discomfort costs
   (e) Inconvenience costs
5. Accident costs
6. Terminal (parking and garaging) costs

B. Potential Benefits

1. User travel benefits
   (a) Perceived user travel benefits
   (b) Non-perceived user travel benefits
2. Facility associated non-user revenues (concession revenue or property taxes)
3. Intergovernmental transfers
4. Other non-user benefits (better view, etc.)
Figure 2-11
Willingness-to-pay as a Measure of Benefits
pansion capacity. Even in the simplest project design problem of a trans-
port nature, there is the problem of existing alternatives and distributed
volumes, or what we can term network effects. In other words, even the
simplest project design problem is not really a mutually exclusive project
design problem.

Assume for the moment, for example, that there is already an
existing facility which is carrying volume from point A to point B which
is congested. The alternatives available to relieve that congestion in-
clude:

(1) improving the existing alternative
(2) expanding its capacity
(3) adding new capacity to carry flow from A to B.

Case (1) and (2) resemble the capacity expansion problem. But
once we consider (3), the problem actually becomes a network problem,
albeit a simple one, since very rarely will the original alternative be-
 tween A and B be abandoned. With both projects—the existing and new
facilities—in operation, the flow between A and B is split (or distributed)
between the two facilities. Thus, as soon as we consider adding a new
facility to carry the demand for an already existing link, we have changed
the problem from a pure project design problem to one which includes net-
work effects.

If this is the case, we need to determine first, what is the
aggregate supply between A and B, and second, how traffic will distribute
itself over the network. The aggregate supply is simply the total capacity
between A and B and is defined as the horizontal sum of the individual
facility supply curves shown in Figure 2-12.

Equilibrium is then determined by equating the demand function with the aggregate supply function.Volumes must still distribute themselves between the two alternatives. We use Wardrop's second principle [44] which states that flows distribute themselves between competing capacities such that prices (or costs) perceived by the users of both facilities are equivalent. With this assumption and the equilibrium price computed, it is a simple matter to select the distributed volumes on each separate link contained in the aggregate supply. Thus, equilibrium volumes are shown as $v_1$ and $v_2$ for each of the two facilities.

2.2.2 Extension to the Network or "Program Selection" Problem

The introduction of an aggregate supply function of the previous section leads us directly and naturally to the more complex, but more interesting problem of network time-staging or program selection. Links or projects are part of a larger system and, in general, cannot be assumed to be independent entities. Introducing an expanded capacity in one part of the network can have effects on flow volumes on adjoining links which will affect additional adjoining links and so on, as we briefly outlined in Chapter 1.0. Additionally, capacity can be added in a number of different ways. The two most common are:

1) expansion of the capacity of an existing facility; the capacity expansion problem

2) adding a link between two nodes of network where no link previously existed; the link addition problem

85
Figure 2-12

Forming the Aggregate Supply Function for Two Parallel Links
A program is defined as a set of projects (capacity expansions and/or link additions) that are to be selected from a large number of projects. In the transport case, both the capacity expansion and link addition problems are faced with the same "systems effects" problem of transportation networks. The approach taken to predict equilibrium volumes is similar in principle to the formation of aggregate supply functions of the previous sections. In practice, equilibrium is accomplished using an iterative procedure between "path" supply functions between nodes and the demand between nodes, known as the assignment procedure in transport planning [45,46].

Ideally, it is a simple matter to test alternative sets of changes to the transport system, and to compare and choose an investment pattern. Practically speaking, however, there are severe combinatorial problems associated with such a simple approach for the following reasons: first of all, the simulation of each state of the system (or each network alternative) results in a separate "equilibrium" for the system; a separate set of link flows, link costs and benefits unique to that alternative for a particular point in time. To be able to make any kind of commitment to a specific resource allocation pattern, each separate alternative should be tested (or simulated and evaluated) followed by a pair-wise comparison with each of the other alternatives. For a set of n potential investments, this can be a formidable task since n individual link additions results in \(2^n\) different programs for comparison. For example, if we assume that there are only four potential link additions to an existing network, there will be 16 distinct plans to simulate \(2^4=16\) representing all the possible
combinations of adding those links; for a mere set of 20 improvements (n=20), the number of distinct plans explodes to 1,048,576. Clearly, this number of plans would be impossible to exhaustively elaborate, even with current computer capabilities.\footnote{\footnotetext{Fortunately, many alternatives can be dismissed with little effort. However, even dismissing the obviously bad plans still leaves a significant number of plans to be tested.}}

Secondly, because of the multistage nature of investments, the simulation of the impacts of alternative network programs must be related to future investment programs. This implies the number of alternatives is multiplied considerably over the static case. We must now simulate each alternative program for each time period in the model.

And finally, if there are also long-run supply-demand dependencies, where the demand functions are affected by the sequence of capacity expansion, we are forced to evaluate alternative programs at each time period for as many ways as there are paths to reach the same program. In other words, it is not enough to evaluate all programs at each period since there are many sequences of investment to reach the network corresponding to a particular plan at a particular period, and each sequence affects the way in which the activity system changes, differently. Hence, the demand for each network is different. This in turn affects the equilibrium volume and the value of that investment pattern.

2.2.3 Uncertainty

The final element of our specific problem statement is that decisions are rarely taken with full knowledge of future outcomes. Generally
there is considerable uncertainty over both the operating characteristics of technology (both existing and new) and the demand for that technology. In most planning approaches, uncertainty is generally ignored and demand is often given as a single point estimate much like the simple capacity expansion models of the previous sections.

Within this study, although we recognize that there are many uncertainties of supply as well as demand, we have chosen to focus on uncertainty of the short-run demand function. In terms of the supply-demand paradigm of Section 2.1, demand is now represented by a number of downward sloping functions, each with a certain probability of occurrence (see Figure 2-13).

Thus, on the one hand, we are introducing a more realistic approach to investment planning by explicitly representing the uncertainty of the future. On the other, we have dealt the final crushing blow to the general equilibrium framework as a multistage planning model in its present form. For each program and each time period, we must also simulate over multiple demand levels.

---

1We will make no distinction throughout this study between the terms risk and uncertainty. As we indicate in Chapter 6.0, we consider the subjective probability estimation procedure of the Bayesian approach to be a reasonable and valid approach to transform true uncertainty into risk. Nonetheless, there are a number of additional approaches that have evolved to treat true uncertainty problems. See Morris [47] for an introductory discussion.
Figure 2-13
Uncertainty Represented by Multiple Demand Functions
2.3 The Sequential Decision Framework

2.3.1 Definitions and Terminology

Before discussing the application of the general sequential decision model to transport planning, we must first introduce some terms used throughout this study in order to define what we mean by an action and a decision tree and the connotation for transport decisions. A decision tree is defined as a sequence of actions, $\Lambda_m$, and state variables, $\phi_k$. Actions are controllable by the planner. State variables are uncontrollable states of nature which may depend on the action chosen and/or some random phenomena of the environment. In the transport case, the alternative actions are the alternative project investments (e.g., link additions to an existing network). The alternative state variable levels results in events, or uncertain outcomes of that investment decision—in this case, we assume the events are alternative demand patterns. Figure 2-14(a) shows a simple, single-stage decision tree, consisting of a set of actions followed by a set of uncertain outcomes. The single stage problem is sometimes referred to as the static problem.

Since transportation investment alternatives are generally very durable goods, they very often have substantial future consequences. A more realistic representation of the transport investment problem would therefore include more than a single stage. The choice of immediate actions is thus influenced by future actions, and the simple diagram in Figure 2-14(b) must be expanded to include multiple actions for different time periods much as we described in previous sections. This is generally
Actions  State Variables

(a)

A Single Stage Decision Tree

(b)

Stage: 1  2  3

A Multistage Transportation Investment Decision Tree

Figure 2-14
Single and Multistage Decision Trees
defined as a multistage¹ decision tree and represents what is commonly known as the dynamic investment problem.

If we represent each action, \( \Lambda_m \), at a particular point in time, \( t \), by a branch of this tree, each action \( \Lambda_m \) now becomes \( \Lambda_{mt} \). Each future action emanating from this branch, \( \Lambda_{j,t+1}, \Lambda_{j+1,t+1}, \text{etc.} \), must include the former action.² Nodes within this tree are defined as either choice nodes or chance nodes. Choice nodes imply a decision is to be made if this point in the tree is encountered. On the other hand, outcomes occurring at chance nodes are determined by some unknown random process. A terminal node is defined as the final chance node beyond which no further action is taken, or alternatively, the node at which the normal computational procedure ceases to operate.

Finally, outcomes or events for the multistage tree can be deterministic or stochastic, and dependent or independent of previous period results. For example, a deterministic tree will have a single outcome branch; for each action, there is only one possible outcome. A stochastic tree has multiple outcomes, each occurring with an associated probability. In addition, the probability of a particular outcome occurring at a stage can be considered independent of the outcome at the previous stage or may be a function of the events which actually occurred at

---

¹Although one must be careful not to imply multistage decision models, always refer to time-related decisions; in some cases, time may not be an essential element; however, in most cases, one does imply the other. Throughout this work, multistage will refer to multitime period decision problems.

²In other words, once an investment takes place, we disallow disinvestment. Theoretically, it is possible for this case to occur—in the transport case, it is very unlikely.
previous stages. (In the most general case, they may be a function
of the previous n stages.) A sequence of a particular series of state
variables occurring is defined as a history; a history of identical
state variables occurring is termed a spine — i.e. a series of low
(high) demand events over time is a low (high) demand spine.

Thus, deterministic situations are ones in which outcomes
associated with an investment are (or are assumed to be) known for
all stages i.e. they are single valued for any given stage (usually
the mean state variable spine). The solution can be given for all
periods and is therefore a sequence of investment decisions.

If uncertainty exists, however — if the problem is of a
stochastic nature — the multistage decision problem can be of two
distinct and quite different forms. These different problem types
can perhaps best be defined by introducing the notion of a sequential
decision process. A stochastic sequential decision process is de-
"a problem which involves the making of two or more
decisions at different points in time, and which has
the property that the later decision(s) may be in-
fluenced not only by previous decisions, but also
by some stochastic parameters whose values will
actually have been observed before later decisions
are made."
In the transport case, it represents a series of sequential, time-staged conditional transportation investment decisions.

The second type of multistage stochastic investment problem is defined as a non-sequential stochastic problem. That is, although the future is perhaps represented as a series of uncertain outcomes, because of other constraints, all decisions (implying the investment sequence) must be made simultaneously, rather than in a sequential manner - all decisions for all time periods are to be made now. The result again is an unconditional sequence of investments as in the deterministic case, but it now recognizes not only the mean demand levels but the extreme values as well. In this case, it makes no difference if the results at the end of the first period are different from those that were expected, since a revised investment sequence is either not allowed, or not expected to occur. Once a decision is made on a particular plan sequence, it is considered to be committed for all time. The procedure therefore decides on the best sequence of investments using an unconditional expected value viewpoint.

Both of these problem types usually assume that the probability distribution of the state variable, though it may be changing over time, is a known distribution. There is another form of uncertainty we will simply introduce at this point for completeness and return to in Chapter 6.0. It involves relaxing the assumption that we know for certain the true underlying probability distribution, given that it is a random process. This can be represented by introducing
uncertainty over the parameters of the probability distribution itself. This we define as the (learning) adaptive sequential decision problem (see Section 2.3.2). At the end of the first period, we not only observe the results in order to make a second period decision, but we use those results to update our knowledge about the distribution itself.

The definitions and terminology we have just described can be represented in terms of a taxonomy of investment problems which are shown in Figure 2-15. Because investment analysis generally implies substantial future impacts and has little to do with single time periods, that portion of the tree can be forthrightly dismissed and we can concentrate on the multiperiod aspects of investment decisions.

In this study, while we will be primarily concerned with the two types of problems represented in Figure 2-15 as stochastic sequential decision processes with distributions both known (or assumed to be known) and unknown, we will refer from time to time to the other problems in this figure, either for comparison purposes or as approximating procedures to the more complex problems.

---

1 See Hadley [49], for a further discussion of these problem types.
Figure 2-15
A Taxonomy of Investment Problems

97
2.3.2 The Definition of Strategies and Sequential Decisions in an Adaptive Sense

The introduction of uncertainty to transport planning, while more realistic in terms of how we view the world, can only lead to additional computational difficulties in terms of planning models as we shall see in Chapter 3.0. Additionally, however, incorporating uncertainty into the model can have two different effects depending on the choice mechanism. First of all, assume for example that although there will be random fluctuations over demands, that all decisions present and future, must be made now. The decision will still represent the complete sequence of decisions over time, which should be different, in general, from the deterministic case. Secondly, suppose that we do not constrain the decision-maker to make all his decisions now. In fact, he need not make them all now. Only the initial decision is irrevocable. All other period decisions can wait until that period comes about. In this way, the decision-maker can make use of additional information that comes his way in the meantime. For example, if he knew, a priori, that demand (which is random) was to be below the mean for 10 years, he would probably shift his initial decision to a smaller capacity facility in the initial period. Therefore, without a priori knowledge, it will be to his benefit to wait for future periods to make future decisions.

Paradoxically, although we are concerned primarily with the initial decisions, we cannot ignore possible future decisions, since first period decisions (and this is especially true for the transport case) have long term future consequences. The solution, of course,
is not to assume certainty and to choose the best sequence of investments, and from this, after implementing the first period solution, to reanalyze the problem. Nor is it to simply choose the best initial investment for stochastic conditions in the first period, ignoring all future possibilities. Rather, the answer is to develop a "tree of decisions", a set of conditional decisions, or what Von Neumann and Morgenstern [50] have labelled, a strategy, as we have defined it in the previous section.

For transport planning, this implies we now develop a series of strategic plans which are then compared for the optimal capacity expansion "strategy". Each plan has an initial investment with each subsequent period having a set of investments, each being optimal for the second stage for a particular demand level which occurs in the first period. Thus, plan A might involve a two-lane facility in the first period, an expansion to 4 lanes in the second if demand is high, or no expansion or the addition of only marginal improvements such as signals or interchanges if demand is low. In the most general case, the initial decision of any strategy will be different from the initial decision of a non-sequential solution, and different still from the initial investment of the optimal deterministic sequence.

The use of the sequential decision approach which results in a strategic solution (in the conditional decision sense) is defined by a branch of operations research known as control theory as an adaptive solution.¹ There have been two different definitions recog-

¹See Hurst [51] for a preliminary discussion of these concepts; see Jacobs [52] for a further discussion.
nized within this body of theory. The first is defined as tracking adaptive; it is also known as signal adaptive. As the name implies, this form of process tracks a signal over time - in the transport case, it is an unknown demand, i.e. the signal is demand response to investments. The use of a tracking adaptive approach is generally implied when we are dealing with a stochastic sequential decision problem.¹

The second kind of adaptation is termed learning adaptive; it is also known as parameter adaptive. It arises when there is uncertainty, not only over demand in the form of a probability distribution, but when the probability distribution itself (in terms of its parameters) is to some degree uncertain. As the process unfolds in time, we not only conditionally choose the investment pattern based on previous outcomes, but we also learn about the true underlying distribution. For example, if demand is observed to be low for some period of time (lower than our probability distribution would seem to indicate was a very likely history) we would more than likely revise our initial estimate of the probability distribution, which in turn will alter the investment strategy subsequent to this point. In some cases, we might expect the learning adaptive solution to choose a non-optimal alternative in order that we might learn more about the true demand distribution, and subsequently make better in-

¹In some problem structures, the sequential and non-sequential solutions are identical as we shall see in Chapter 5.0, i.e. the best strategy is an unconditional sequence.
vestment decisions.

We explore these concepts further in the following chapters; tracking adaptive in Chapter 4.0 and the learning adaptive model in Chapter 6.0. We now turn to a brief review of the literature in terms of existing approaches to time-staging of transportation investments.

2.4 Survey of the Literature

There have been a number of recent notable attempts to introduce time and staging analysis into the investment planning framework. Although our primary interest is in the area of transport investment planning and the application of uncertainty and the sequential decision framework, we will not limit our review to either transportation or time-staging models and algorithms. There are a number of approaches not developed in the transport field nor directly incorporating time in a meaningful way which are therefore included because of their current or expected future relevance to the sequential decision framework.

A general survey of the literature has shown that all of the current approaches to time-staging of transport investment fall naturally into the two major categories of 1) project design and 2) program selection. Within each of these categories, most of the emphasis has been on either applying prescriptive optimization algorithms to fairly simplified problem structures or on massive complex simulation models which have no concern for the search problem of a combinatorially large solution space. The one exception has been the recent application of discrete optimization techniques of branch and bound and dynamic programming to transport investment network flow problems. Even these
approaches however, require some extremely strong assumptions about the solution space.

The first major distinction between time-staging approaches is therefore based on whether or not we are dealing with single or multiple projects, i.e. project design or program selection. Program selection is further divided into three categories which become progressively more complex:

1) Capital Budgeting Models
2) Network Flow Models
3) Activity Growth Models

Capital Budgeting, which has built up a considerable literature in the field of management science primarily, generally assumes all benefits are exogenously specified and single-valued. The exercise is usually one in combinational mathematics of selecting the best sequence of investments subject to capital (budget) constraints. Network flow models have only recently extended the capital budgeting literature and embedded a network flow model which can account for network dependencies and internally predicts the flows, and hence, the costs and benefits. They have been generally limited to linear flow models, however. Additionally, they usually ignore the dependency of supply-demand. Recently, developments in branch and bound techniques have placed fewer constraints on the form of the flow model. And finally, growth models, the most complex, add the constraints of long-run supply-demand dependencies in addition to normally using a complicated network flow simulation procedure.

102
Thus, referring to Chapter 1.0's brief outline of interdependencies, project design has traditionally ignored most dependencies until recently; capital budgeting focuses on financial dependencies; network flow models add project, supply-demand, and financial dependencies; and activity growth models capture true timing interdependencies as well.¹

It can generally be assumed that as the model proceeds from project to program selection and within program selection, from capital budgeting through growth models, the procedure becomes considerably more complex but more closely models the real world. Thus, as we proceed through this taxonomy, the numbers of alternative investments that can be investigated becomes substantially smaller for a given computational budget. It is therefore understandable that most of the research effort has been in the project design and capital budgeting area. Notably, most of the tractable solution procedures (calculus, mathematical programming, etc.) are most useful for the project design and capital budgeting problems. Similarly, it is in these simpler problems that most of the work concerning uncertain payoffs has been attempted. The following is a review of the work which was most relevant to our study, or which appears to be potentially useful for extensions to this study.

¹This is not entirely true. For example, a number of capital budgeting models have included project dependencies. Additionally, it is not clear whether some network flow models should be classed as a capital budgeting model or as a network flow model with added budget constraints. Clearly, the dividing lines between these categories is not well-defined. Nonetheless, over all, this appears to be a good way to categorize the field in terms of time-staging models. For a different scheme, see Burns [53], and Kuhn [54].
Project Design

The project design literature relevant to this work has been referred to earlier; Chenery [55], Manne [56], Muhich [57], have all been concerned with capacity expansion (or time-staging) models, dealing primarily with the tradeoffs between economies of scale and opportunity losses due to excess capacity, for pipeline and water treatment plant expansion. Manne and Muhich introduce uncertainty and produce a number of interesting results, such as suggesting that we lower the discount rate as a surrogate for the capacity expansion problem under uncertainty, which implies increasing the capacity over the deterministic solution. We will return to discuss this result and compare it with the results of our model in Chapter 5.0. Russel [58] also develops a deterministic capacity expansion model for water treatment plan expansions similar to the former models. The common feature of all these approaches has been the emphasis on the timing and scale of expansion using fairly tractable mathematical models. Chenery and Muhich use calculus; Manne suggests dynamic programming, and Russel uses two techniques; a non-linear programming formulation and a direct search technique. Manne and Muhich have assumed uncertainty to be represented by a stationary and known probability distribution. Solutions are therefore given as unconditional investment sequences. Russel only briefly explores the effects of uncertainty through questions of the type "what happens if demand is different from the expected...".

Turning to specific transport applications, there is a considerable literature on project design, most of it dealing with the determinis-
tic problem. We mention the most relevant here: Wohl and Martin [59] deal with project design recognizing future impacts of immediate investments but essentially ignore the timing question. Tarplay and Drake [60], Thygeson [61] and Marglin [62] (though not dealing with transport problems, Marglin clearly fits in with this group) all develop fairly simple timing models which recognize that substantial benefits can be achieved by considering postponement of investment, i.e. the timing dimension as well. Marglin is especially insightful in recognizing, and describing, the issue of timing. Winfrey [63] is concerned not so much with timing (when should investment occur), as he is with whether or not a full 4 lane investment project intended to be implemented immediately should be staged in a 2 stage construction sequence, and if so, when should the 2nd stage occur. He focuses mainly on the project design staging question given that some capacity is needed now. Thus, he ignores the question of delaying the initial investment, the question of aggregate supply and even supply-demand dependencies. Cole [64], in a fairly abstract study aimed at transport capacity expansions, introduces uncertainty in the form of a probabilis-
tic demand structure but limits his stages to coincide with the economic life of the facility. His solution, therefore, is also a non-sequential (unconditional) investment sequence i.e. he can decide on an investment sequence now, once and for all, even though there is uncertainty over demand. Howard and Nemhauser [65] introduce dynamic programming for the project design problem for the transport case but their treatment is fairly theoretical. However, they do treat uncertainty and supply-demand dependencies which most of the previous works have ignored.
The project design problem is analogous to the inventory control problem [66], and for the most part, it appears that water resources has made the most of this work. In part, the lack of transport application may be due to the difficulty of "breaking out" a project from the rest of the network.

Program Selection

The literature of the capital budgeting field contains perhaps the most references of all four categories, and has perhaps stimulated the most work in the area of program selection for the transport problem; at least for the application of capital budgeting to transportation program selection and network flow models. This is due to a number of factors; undoubtedly, the primary one being the explosive development of mathematical programming techniques for handling large-scale combinatorial problems, and capital budgeting's "perfect fit" with these approaches.

The most relevant works in this area, i.e. ones which deal with multiple project selection over multiple time periods, and are useful for the transport time-staging problem are the following: Marglin [67] (again), in what has to be described as a classic study, appears to be one of the first to deal with the program selection problem, although, he admittedly deals with dependencies caused solely by budget constraints; Weingartner [68] is another classic work, which thoroughly explores mathematical programming and the capital budgeting problem. In a later paper [69], he also presents a survey of a number of techniques useful for producing solutions to the budgeting problem.
Consad [70] also presents a survey (a later one) on capital budgeting models as background for their own efforts in staging transport investments, still in terms of abstract projects (i.e. projects, though intended to be transport projects, are represented solely by a set of costs and benefits). In all of this, they assume away the stochastic aspects of the problem by relying on the Central Limit Theorem and basing their decisions on average payoffs. One of their proposed models for selecting staged sequences of investment for the Northeast corridor project was a quadratic programming model which can handle project dependent costs and benefits quite easily.¹

¹Using the transport investment problem as the basic motivation, a number of other authors also concern themselves with program selection models for transportation using only exogenously specified costs and benefits. Mori [71] explores the use of dynamic programming for multiple project selection and introduces capital budget constraints. He shows the limitations of this technique for more than two or three stages. Meyer and Straszheim [72] also review a significant portion of the capital budgeting literature relevant to transport planning and present a fairly concise and clear treatment of the dual problem, shadow prices, and internal (vs. external) opportunity costs of projects.

¹It is interesting to note that they also apply it in a two phase scheme which first produces a set of good plans for the horizon. They then suggest applying the model over time using the reduced set of plans to produce the optimal staging sequence. See Section 4.2.2.4.
There are a great many other capital budgeting articles, far too numerous to completely identify here. The most relevant appear to be those which introduce uncertainty and the multi-period nature of the problem in some way into their model. Salazar and Sen [73] give a good summary of the techniques for handling uncertainty for the capital budgeting problem. Paraphrasing their work in part, the major approaches to handling uncertainty for the multi-period capital budgeting problem appear to be the following:

(1) Stochastic Linear Programming [74,75,76] - there are two approaches: the first attempts to obtain a distribution of outcomes by repeated solution of a linear programming (LP) problem; the second uses the probability conditions as linear equations and solves for a solution which gives maximum expected value directly.

(2) Linear Programming under Uncertainty [77] - this is an attempt to account for the true sequential nature of decision-making. Essentially, the method is to formulate an LP problem in two stages. Values for first stage decision variables are selected followed by an observation of the random variables. Finally, values for the second-stage decision variables are computed and the optimal initial investment chosen.

(3) Chance-Constrained Programming [78] - the expected value of a functional is maximized subject to probabilistic constraints.

The authors go on to identify a number of notable works in these areas. Finally, they summarize a whole series of approaches which can be closely related to our work (to be discussed in the following chapter): the use of stochastic decision trees to analyze
sequential decision making under uncertainty [79,80,81,82]. They themselves present a decision tree formulation which was inspirational for some of the ideas of this study.

In all of the above capital budgeting models, the primary emphasis for the most part has been on reducing the combinatorial aspects of a program selection problem, recognizing the constraints of a limited budget, but most importantly, with the assumption that the costs and benefits (or cash flow) are exogenously specified and independent of the sequence chosen (if uncertainty is incorporated, then the distribution of costs and benefits is exogenously specified).

We now turn to describe a number of techniques which have obviously been an outgrowth of the concomitant interest shown in capital budgeting and mathematical programming techniques; namely network flow models. In these formulations, the primary characteristic is the fact that cost and benefits are not exogenously specified, fixed quantities, but depend on some prediction mechanism. In some cases, it can be internal (as in linear flow models) and in others, it is a completely separate model. Notably, they all deal with a deterministic and fixed demand structure.

Hershdorfer [83] applied a branch and bound algorithm to the single period, link addition problem developing a linear programming flow model to determine the measure of effectiveness of network changes. Roberts [84] at the same time, but independently, was exploring the use of the same branch and bound algorithm coupled with a heuristic backward-stepping, time-sequencing algorithm for the multi-period problem.
Bergendahl [85], in a similar approach to Roberts, used a linear programming flow model for predicting the flow pattern for any investment plan at a stage, but employed dynamic programming to search for the optimal time-staged investment sequence. Bergendahl also gives a fairly good review of related solution procedures for the deterministic time-staging problem. Taborga [86], also, has employed a dynamic programming algorithm, for the problem of seaport expansion, using a queuing model for producing returns for any specific plan.

Morlok [87] also has proposed a dynamic programming procedure for producing time-staging strategies in the Northeast Corridor context, ignoring the network effects of multiple and overlapping paths.

A second major approach to reduce the combinatorial problem of time-staging models is the use of the discrete optimization technique of integer branch and bound programming. Ochoa [88] and a number of other authors explore the use of branch and bound for capital budgeting problems in transportation. Ochoa and Silva [89] apply a branch and bound algorithm and a branch and backtrack algorithm to the network investment (single period) problem using the traditional assignment model as a flow prediction mechanism. A number of related works (Ichbiah [90], Bivins [91]) explore the use of that branch of mathematical programming which is experiencing a great deal of attention these days; branch and bound and implicit enumeration schemes, for reducing the combinatorial problems of program selection. To date, there has been little done with these techniques for either the
multi-period, or the stochastic problem. Bivins, however, suggests an algorithm based on his work, for extending the single period problem to multi-time periods, after Roberts' formulation.

Additionally, there are a number of heuristic approaches for the network flow problem which concentrate mainly on the link addition or capacity expansion, single period problem when using a simulation model. They are clearly related and should be explored further for their relevance to the time-staging problem: Barbier [92], Stairs [93], and Spencer [94] all discuss ways to select investment plans for testing in a simulation model, which correspond to a form of direct search procedure (see Section 3.3.2). Bhatt [95] in a thesis related to the same larger study effort that this study is related to, also explores the use of heuristics for the single period investment problem which should prove useful for the time-staging problem.

In another series of related work, Lai and Schaaake [96], and Lai [97], explore the investment problem for water resource pipe networks for deterministic, and stochastic conditions, respectively.

And, finally, we turn to the area of activity growth models. Although this area would seem to have the most impact on actual investments of a transport nature, there has been surprisingly little work done on the use of such models for evaluating time-staging strategies. The reason for this is obvious, however. It is because of the complexity of the models employed, the detail involved in just running a single simulation at a stage for one investment plan that little has
been done with staging or staging under uncertainty. It is precisely in this area that the value of the sequential decision model and heuristics for reducing the search will be useful. The two most significant studies which employ an activity shift or macro-economic model in conjunction with a transport model are the Northeast Corridor Study (NEC) [98] and the Harvard-Brookings study [99].

Nutter [100] presents a heuristic iterative scheme for producing time-staging strategies using such a model in the NEC context, and the Harvard-Brookings models are used to test different investment time-staged sequences which are exogenously specified.

Although there are a number of works which are also related to our study, such as references on statistical decision theory, and adaptive control, etc. we will delay describing them until the appropriate section where they will be discussed in more detail. A couple deserve mention here for their bibliographies; Murphy [101] compiles a fairly brief but encompassing bibliography on works related to the issues of adaptive investment models, multi-project investment under uncertainty, information theory, and so on. Hillier [102] in the third monograph of a series on Budgeting Interrelated Activities, also presents an extensive bibliography on related material to the investment problem under uncertainty. And finally, Massé [103] presents a fairly comprehensive survey of work done by him and others in the area of planning investment under uncertainty. He also gives an interesting discussion of the history of the notion of a strategy and the sequential decision process.
2.5 Summary and Conclusions

In this chapter, we have explored the nature of the transport investment problem from the point of view of (1) the objectives of transport investment policy and (2) the general methodology that has evolved over the past 10 years or so in parallel with the objectives. After discussing the general problem and presenting the general equilibrium framework, we explored the two major application areas of this framework: project design and program selection. For the most part, each of these two approaches attempt to solve part of the transport problem, assuming away some issues while focusing on those which seem most appropriate to a particular problem context.

Using project design as the basic problem environment (i.e. ignoring multiple project dependencies and network effects), we explored some of the more obvious economic tradeoffs of staging and economies of scale and defined the project design problem for the transport case.

The problem of program selection considers inclusion of additional dependencies of network effects, financial dependencies and so on. Three major types of problems within the area of program selection have evolved; capital budgeting, network flow, and activity growth problems. For the most part, the first two approaches have focused almost exclusively on the combinatorial problem to the exclusion of real world application, while the latter approach is centered on the real world problems of prediction, largely ignoring the problem of search and choice.
We then discussed two of the most obvious extensions to the
general planning framework, that of uncertainty and the sequential
decision framework. The multistage approach of sequential decision
type, particularly with a capability for learning about uncertain
state variable probabilities, gives greater flexibility in planning,
and represents in fact, how decisions are actually made - in a sequential
incremental fashion. We should now be able to take into account un-
certainty of both new technologies, and demands for that technology, in
a more explicit fashion.

And finally, the chapter has presented a basic survey of trans-
port planning models and techniques which have been developed as
partial tools for this general transport planning framework.

Based on the nature of current planning practice, a survey
of the literature, and actual implementation policies, we have con-
cluded that there are at least two major remaining deficiencies in
the general methodology; that of handling uncertainty and the sequential
nature of investments. More specifically, after reviewing a number of
study approaches, we can conclude the following:

(1) the transport investment problem is complex in
theory and depends largely on whose point of view
we are concerned with;

(2) the general problem has been approached from a
variety of different perspectives, each emphasizing a
certain commitment to a profession, to a mode, or a

114
philosophy which emphasizes some aspects and largely ignores others;

(3) a number of computational techniques have been employed in these approaches in an attempt to solve a number of distinct sub-divisions of the planning problem, all of which fall into the two major categories of either project design or program selection;

(4) none of these techniques or algorithms can be used satisfactorially in all of these basic problems of transport investment planning; and finally,

(5) the one major deficiency of all of these techniques and of the transport planning framework is that no one technique now available can satisfy all of the constraints and interdependencies, and still recognize both uncertain future outcomes and the sequential nature of transport investments.

To this point we have presented the basic environment in which transport planning takes place. We have defined some of the specific issues involved and some current approaches to solving these issues. Recognizing the two major deficiencies we have chosen to explore, uncertainty and sequential decisions, we now turn to an explicit mathematical statement of the problem and to explore some alternative solution procedures for this general framework in more detail.
Chapter 3

POSSIBLE SOLUTION TECHNIQUES FOR THE
SEQUENTIAL DECISION MODEL
Chapter 3

POSSIBLE SOLUTION TECHNIQUES FOR THE
SEQUENTIAL DECISION MODEL

In the previous chapter, we presented the traditional framework of transportation systems analysis as well as a summary of some of the more prominent approaches and their shortcomings as examples of current practice in investment planning. The remaining sections of the chapter were then devoted to describing the structure of the sequential decision model and the application of this analysis framework to transport investment planning.

The purpose of this chapter is to first, present a more mathematical formulation of the time-staging problem, and then to describe a number of alternative algorithms as possible solution procedures for the sequential decision model when applied to transport investment planning. A number of potentially useful approaches are presented and discussed; then the procedure which we propose to use is described.

3.1 Description of the Problem

The first two chapters have emphasized our general concern with investment planning models, and specifically, in the uncertainty of future demands and the application of the sequential decision framework to transport planning. In order to focus this interest, we will concentrate on two well-defined and specific transportation investment problems taken from the literature--the problems of project design and program selection. Although these example problems will involve only a single technology
(highways), most of the framework we will develop will be applicable to other investment problems with little loss of usefulness. By focusing on a single mode we will be able to ignore the details (and costs) of more complex simulation procedures, demand models, etc., necessary to capture the effects of both substitutable and complementary modes, and most importantly, to concentrate on the sequential decision framework itself.

The choice variables of these two problems are also limited to the timing, \( t_j \), and scale, \( \Psi_j \), of additions and expansions to an existing transport structure. The net benefits of alternative investments are determined at each stage by first running an iterative and non-analytic simulation model. This produces the equilibrium flows for a specific network alternative which are then used to obtain the gross benefits. From these gross benefits we then subtract the fixed, maintenance, and operating costs of the network in place at that time.

Constraints on these problems can be of three types:

1. Network interdependencies: links are normally topologically connected and a flow distribution rule is required to determine equilibrium flows.

2. Budget limitations: at any stage, there are restrictions on the total amount of capital which can be expended; this constrains how many links can be added to the network at any point in time.

---

1 Relaxing one or more of these constraints makes the solution significantly easier. For the most part, we will ignore constraint number 2, budget limitations, and concentrate on the problems of project design and network program selection with constraints 1 and 3.
3. Supply-demand dependencies:

(i) the flow on any link is limited by the capacity of that link as a short-run constraint;

(ii) dependencies between long-run demand shifts and capacity changes (in terms of accessibility) affect the volumes and the returns for any network.

Additionally, we have incorporated uncertainty into the problem by assuming the parameters of the short-run demand function are random variables and are represented as such by a discrete probability density function. This results in a set of flows, $F(\Lambda_{mt}, \phi_{kt})$, for each network alternative, $\Lambda_{mt}$, at stage $t$, each flow corresponding to a single value of the uncertain parameter set, $\phi_{kt}$.

The objective of the investment planning model we plan to develop in this and the following chapter is to choose the investment pattern which maximizes expected discounted net benefits over a pre-specified planning horizon. The model does this by searching for the best investment strategy from among all strategies (see Section 2.3), using the sequential decision model as the basic framework.

This investment problem is represented again in Figure 3-1 as a decision tree in extensive form. $\Lambda_{mt}$ represents investment alternatives, and $\phi_{kt}$, the set of uncertain demand parameters. For each $[\Lambda_{mt}, \phi_{kt}]$ couple, a simulation model determines the set of equilibrium flows, $F(\Lambda_{mt}, \phi_{kt})$ for any stage $t$. What is not shown are the constraints on passing from one stage to the next, the exact form of the simulation model, and so on. We now turn to a detailed formal description of this tree for the transport investment case.
Figure 3-1
The Transport Investment Problem Represented in Extensive Form
3.2 Mathematical Formulation of the Problem

3.2.1 Link Decision Variables

Since our concern is with the long-run investment problem, we define the primary decision variable to be in terms of link investments or improvements. Let \( \ell_j \) represent the decision variable for link \( j \), which can take on the possible values 0 or 1, representing the two possible cases of no investment and investment occurring. The vector of decision variables, \( \Lambda_m \), is the set of \( \ell_j \)'s specifying those links which have already been constructed and those which are still potential candidates for investment. Then

\[
\Lambda_m = [\ell_1, \ell_2, \ell_3, \ell_4, \ldots, \ell_n]
\]

and \( \ell_j = 1 \) if link \( j \) has been implemented, \( \ell_j = 0 \) otherwise. For the network shown in Figure 3-2 with 11 links, the vector

\[
\Lambda_m = [11011100001]
\]

represents links 1, 2, 4, 5, 6, and 11 being in place with 3, 7, 8, 9, and 10 remaining as potential investment alternatives. In general, for a network with \( n \) links, there are \( 2^n \) such network plans or states, \( \Lambda_m \), representing all possible combinations of \((0,1)\)'s for each link in the set \( \ell_n \), as shown in Table 3-1.

3.2.2 Investment Programs

Using the full set of network plans \([\Lambda]\), we can now define the set of investment programs, \( \xi \), the set of network changes necessary to go from one plan to another, as follows:

121
Figure 3-2

Alternative Network Plan $\Lambda_m$ Describing Current and Potential Link Investments
## Table 3-1
Potential Network Plans for a Set of n Links

<table>
<thead>
<tr>
<th>Links $l_j$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>......</th>
<th>(n-1)</th>
<th>(n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network Plans $\Lambda_m$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>......</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>......</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>......</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>......</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>......</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\vdots$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\vdots$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\vdots$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(2^n)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
\[ \xi = \bigcup_{j} \Lambda_{i} \Lambda_{j} \] for all combinations \((i,j)\), where \(\Lambda_{i} \supseteq \Lambda_{j}\)

\[ i, j = 1, \ldots, n. \]

Thus, the set of investment programs is determined by the difference between plans \(\Lambda_{j}\) and \(\Lambda_{i}\), providing \(\Lambda_{i}\) has at least all the same positive elements of \(\Lambda_{j}\). In other words, the sum of the elements of \(\Lambda_{i}\) must be greater than or equal to the sum of the \(\Lambda_{j}\) elements and be coincident for the plan with the fewer elements. This merely ensures that a program is a positive one of investment (no disinvestment allowed).

The full set of programs, \(\xi\), is composed of two parts: the first \(2^n\) programs, \(\xi_0\), are identical to the first \(2^n\) plans, \([\Lambda_{2^n}]\), and represent the investment changes from the null network plan, \(\Lambda_0 = [000\ldots0]\), to each plan of the \([\Lambda_{2^n}]\) set. The second set, \(\xi_{-1}\), is generated using the above equation by considering investment from one plan to another, if permissible.

Adding the dimension of time to the decision variables \(\ell_{j}\), we define \(\ell_{j}(t)\) or \(\ell_{jt}\) as representing that link being implemented or not at time period \(t\). Similarly, all \(\Lambda_{m}\) now become \(\Lambda_{mt}\), and \(\xi_{m}\) become \(\xi_{mt}\). This has no effect on the total number of physical investment programs \(\xi\); but it does increase the number to be evaluated since the same program in two different periods results in two different sets of consequences, because of changing population, demands, and accessibility structure (see Section 5.2.3).

3.2.3 **Link Investment Matrix**

Once a particular link decision variable \(\ell_{jt}\) assumes an integer value 1, all subsequent elements along the time dimension are equivalent:
$\ell_{jt'} = 1$ for $t' > t$. To fully specify a sequence of investments in the time dimension, we need to specify a two dimensional array $[\Lambda_{it}]$, $i = 1, \ldots, n$, $t = 1, \ldots, N$, shown in Table 3-2. This we will label the link investment matrix. We define a particular first stage investment of the sequence as $\sigma_i = \Lambda_{i1}$.

### Table 3-2

The Link Investment Matrix

<table>
<thead>
<tr>
<th>Time Period t</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>\ldots</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project k</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>\ldots</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>\ldots</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

$[\Lambda_{it}]_{t=0} = \text{existing network}$

$\Lambda_{i1}, \Lambda_{i1+1}, \ldots, \Lambda_{i5}, t$
Any particular plan $\Lambda_{it}$ at stage $t$ describing the links in place, is fully represented by a column vector in this matrix. Any row fully specifies when a project was first implemented. Note, however, that the complete matrix $[\Lambda_{it}]$ only describes a single time-staged investment sequence. A complete enumeration of all possible sequences is difficult to describe explicitly in matrix form—-it results in all possible combinations of permissible plan sequences for all time periods.

For ease of notation, and for programming purposes, we have also used a compact notation for representing an investment vector or sequence. Let $S_t$ be the compact notation for the link investment matrix, $[\Lambda_{it}]$, that consists of a single vector which contains the time of implementation for each link as shown in Table 3-3.

<table>
<thead>
<tr>
<th>Project $l_k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Investment Vector ($S_t$), time of implementation</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

### 3.2.4 State Transformation Matrix

The number of permissible plans is less than $2^n$ at each possible stage for any given previous plan. For example, certain links which are constructed in time period $t-1$ will preclude alternatives at time $t$ (all those plans, $\Lambda_{mt}$, which do not have the set of investments at $t-1$).
It is useful for computational purposes to structure the permissible transformations from one plan to another in what is labelled a state transformation matrix, $T^\Lambda$, as represented in Table 3-4. This matrix thus represents whether an investment program, $\xi_i$, is allowable at any stage, given the network plan at the previous stage.

**Table 3-4**

A Typical State Transformation Matrix, $T^\Lambda$

<table>
<thead>
<tr>
<th>Plan $\Lambda_{i,t}$</th>
<th>1 2 3 4 5 6 7 .... $2^n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plan $\Lambda_{i,t-1}$</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1 1 1 0 0 1 1</td>
</tr>
<tr>
<td>2</td>
<td>0 1 1 0 0 1 1</td>
</tr>
<tr>
<td>3</td>
<td>0 0 0 1 1 0 0</td>
</tr>
<tr>
<td>4</td>
<td>0 0 0 0 1 1 0</td>
</tr>
<tr>
<td>5</td>
<td>0 1 1 0 1 0 0</td>
</tr>
<tr>
<td>6</td>
<td>1 1 0 0 0 1 1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$2^n$</td>
<td></td>
</tr>
</tbody>
</table>

3.2.5 **Equilibrium Flows at any Stage**

The equilibrium flows for any network plan $\Lambda_{mt}$ at time $t$ are determined by equating the short-run supply and demand functions for a given socio-economic conditions. Supply is represented by a series of link supply functions, as described in Section 2.2 and shown in Figure 127.
3-3(a), one function for each type or scale of facility. These are then aggregated in various ways to form total path supply functions for the network "supply" between zones i and j. (We will discuss this further in Chapter 5.0).

Demands are represented as interzonal (supply-sensitive) functions, again similar to those presented in Section 2.2, for given levels of population, income, etc., between zones i and j, as

\[ V_{ij} = \alpha_0 (P_{i}^{P_{j}}) (Y_{i}) (E_{j}) \ldots (c_{ij}) \]

where

- \( V_{ij} \) = the interzonal demand between zones i and j
- \( P_{i}^{P_{j}} \) = populations of zones i and j
- \( Y_{i} \) = average per capita income of zone i
- \( E_{j} \) = employment in zone j
- \( c_{ij} \) = the total perceived price to the user for a trip between i and j
- \( \alpha_0, \alpha_1, \alpha_2, \alpha_3, \ldots, \beta \) = parameters of the demand function

A typical demand function is shown in Figure 3-3(b).

Equating these functions in an iterative, incremental fashion produces a set of flow, \( F(\lambda_{mt}, \phi_{kt}) \) and in turn, a set of benefits \( G(\lambda_{mt}, \phi_{kt}) \) for any specific state variable set, \( \phi_{kt} = [\alpha, \beta] \) (Figure 3-3 (c)).

The one final feature of our model is the fact that we assume that the state variables, \( \phi_{kt} \), the set of parameters of the demand function, are not known, deterministic quantities but are subject to some degree of
Figure 3-3
Determining Equilibrium Flows
129
uncertainty. This uncertainty is represented by a discrete probability distribution, \( p_k(\phi_{kt}) \). The effect of this uncertainty is shown in Figure 3-4 for variations in both \( \alpha_0 \) and \( \beta \) with the other parameters held constant.

3.2.6 The Objective Function

For an infinite planning horizon, the investment criterion for both the project design and network program selection problems can now be stated explicitly as follows: choose that time-stage strategy of investment which produces maximum expected net benefits, \( \bar{B}(\sigma_i) \), and a first stage investment, \( \sigma_i \),

\[
\bar{B}(\sigma_i) = \Sigma \left( \sum_{t=1}^{\infty} G(m_t, \phi_{kt}) \cdot p_k(\phi_{kt}) - (M(m_t) + K(M_{mt})) \right) (1+p)^{-t}
\]

where \( G(m_t, \phi_{kt}) \) represents the net benefits from any flow, \( F(m_t, \phi_{kt}) \), determined through an incremental non-analytic simulation procedure and is a random variable which, in turn, depends on the random variable demand parameter set, \( \phi_{kt} \). Therefore, it must be multiplied by the appropriate probability, \( p_k(\phi_{kt}) \), and summed over all possible levels of the parameter \( k \). The second set of terms represents the maintenance, \( M(m_t) \), and capital cost, \( K(M_{mt}) \), resulting from investment program \( M_{mt} \) in effect for any given time period \( t \). These terms are discounted using the discount factor \( (1+p)^{-t} \) and summed over all possible periods, \( t = 1, \ldots, \infty \). Clearly, for computational reasons, we need to reduce the number of simulation periods to something less than infinity. Therefore, we convert the model to a pseudo-fixed horizon model by limiting the simulation periods to \( \hat{N} \). The
Figure 3-4
The Effects of Demand Parameter Variations
131
future beyond \( \hat{N} \) to \( \infty \) is accounted for by a terminal evaluation function, \( \Gamma_k(\Lambda_{mN'}, \phi_{k'}) \). Therefore, the expression for \( \bar{B}(\sigma_i) \) becomes

\[
\bar{B}(\sigma_i) = \sum_{t=1}^{\hat{N}} \left\{ \sum_{j} \left[ (\xi(\Lambda_{mt}, \phi_{kt}) \cdot p_k(\phi_{kt})) - (M(\Lambda_{mt}) + K(\xi_{mt})) \right] (1+\rho)^{-t} \right\} + (1+\rho)^{-\hat{N}} \Gamma_k(\Lambda_{mN'}, \phi_{k'}).
\]

Because of the unique nature of our model, the exact form of these terminal correction factors, \( \Gamma_k(\Lambda_{mN'}, \phi_{k'}) \), will vary, as does the specific location of \( \hat{N} \). This will be discussed in more detail in Section 4.2. Each transport investment plan, \( \Lambda_{mt} \), of the full set of \( 2^N \) plans designed for some future horizon year generates a time-stream of primary benefits and costs which is a function of the sequencing policy for that plan. For each plan there is generally a number of combinations of time-staging strategies to reach that plan, each with a different set of consequences, as we showed previously.

The purpose of the planning model then, is to compute the expected net present value for various combinations of investment programs, \( \xi \). Each possible program can only reveal its impacts through a complex non-analytic simulation program and must be tested explicitly. Because of the large number of possible programs, we need therefore to be fairly selective in the programs we choose to test.

3.2.7 The Combinatorial Effects of Time-Staged Investment Sequences

Application of the extensive form of the decision tree to either of the two problems we have selected is justified because of the complexity
of the tree and the violation, in most cases, of the primary assumptions
of other solution techniques such as linear or dynamic programming, as we
shall see in Section 3.3.1. Although the extensive form does allow a more
general model formulation and requires fewer assumptions, it too can pro-
duce extreme combinatorial problems by its very structure. The purpose
of this section is to explore the specific nature of this combinatorial
problem.

Assume for the moment that we are considering n potential link
investments which results in $2^n$ plans $\Lambda$ that can occur at any stage and
that there are j stages in all. Budget constraints on capital investment
are further assumed to limit the total number of projects that can be
selected of these n to m for the currently selected horizon, N. The total
number of distinct plans, $\Lambda_{i}$, that can be selected for the horizon period
is then given by the number of ways of taking m distinct objects from a
total of n,

$$i_{\text{max}} = \frac{n!}{(n-m)! \cdot m!}$$

Next, assuming investment is allocated sequentially to each of the j
periods with a maximum of k links constructed per period, we can also
determine how many alternative investment sequences there are to reach
these $i_{\text{max}}$ plans. Using the same procedure we used to determine the total
number of horizon plans, we see that there are

$$\frac{n!}{(n-k)! \cdot k!}$$

combinations of first period plans.
Given these plans, the total number of second period plans are reduced by k first period links, or, there are

\[
\frac{(n-k)!}{((n-k)-k)! \cdot k!}
\]

combinations of second period plans.

Continuing in this manner to the \(j^{th}\) period, the number of \(j^{th}\) period plans is given by

\[
\frac{(n-(j-1)k)!}{((n-(j-1)k)-k)! \cdot k!}.
\]

Finally, assuming all remaining projects are assigned to a dummy \(j+1^{st}\) period, the \(j+1^{st}\) period set of plans number

\[
\frac{(n-jk)!}{((n-jk)-k)! \cdot (n-jk)!}
\]

Accumulating the total number of plans for each of the \(j\) stages, the total number of investment sequences \(S_{n,k,j}\), is finally given by

\[
S_{n,k,j} = \frac{n!}{(n-k)! \cdot k!} \cdot \frac{(n-k)!}{((n-k)-k)! \cdot k!} \cdots \\
\cdots \frac{(n-(j-1)k)!}{((n-(j-1)k)-k)! \cdot k!} \cdot \frac{(n-jk)!}{((n-jk)-(n-jk)! \cdot (n-jk)!}
\]

\[
= \frac{n!}{(k!)^{j} \cdot (n-(k\cdot j))!}.
\]

The effect of the magnitude of these numbers when considering simulation models is quite dramatic for even a small number of potential investment alternatives \(n\), and time periods \(j\). For example, for \(n = 7\)
potential link investments, and \( m = 4 \) horizon links in place (limited by the budget), the total number of plans possible at the horizon is

\[
I_{\text{max}} = \frac{7!}{(7-4)! \cdot 4!} = \frac{7!}{6} = 35.
\]

Assuming \( j = 2 \) periods and \( k = 2 \) links per period, the number of complete sequences possible is given by

\[
S_{7,2,2} = \frac{n!}{(k^j)(n-(k\cdot j))!} = \frac{7!}{(2)^2(7-(2\cdot 2))!} = \frac{7!}{(2)^2(7-4)!} = 210.
\]

This means that there are 35 different horizon plans and 210 different ways of reaching those 35 end-states.

By keeping all else constant, but merely increasing the number of link alternatives to \( n = 11 \), the number of investment sequences jumps to 7920. Table 3-5 shows the effect of sequencing for other values of \( n, k, j \).

**Table 3-5**

The Combinatorial Effects of Multi-Period Models

<table>
<thead>
<tr>
<th>Number of Potential Link Investments ( n )</th>
<th>( S_{n,k,j} ), The Number of Time-Staging Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of Periods ( j )</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Maximum number of investments per period</td>
<td></td>
</tr>
<tr>
<td>( k=1 )</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td></td>
</tr>
<tr>
<td>110</td>
<td></td>
</tr>
<tr>
<td>380</td>
<td></td>
</tr>
<tr>
<td>Maximum number of investments per period</td>
<td></td>
</tr>
<tr>
<td>( k=2 )</td>
<td></td>
</tr>
<tr>
<td>210</td>
<td></td>
</tr>
<tr>
<td>7,920</td>
<td></td>
</tr>
<tr>
<td>28,310</td>
<td></td>
</tr>
<tr>
<td>Maximum number of investments per period</td>
<td></td>
</tr>
<tr>
<td>( k=1 )</td>
<td></td>
</tr>
<tr>
<td>N.A.</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
</tr>
<tr>
<td>( \text{N.A.} )</td>
<td></td>
</tr>
<tr>
<td>( \text{N.A.} )</td>
<td></td>
</tr>
<tr>
<td>( &gt; 5 \times 10^{11} )</td>
<td></td>
</tr>
<tr>
<td>( &gt; 1 \times 10^{14} )</td>
<td></td>
</tr>
</tbody>
</table>

135
The example demonstrates dramatically the combinatorial problems of a multi-period time-staging model. For the simple example of \( n = 7, j = 2, k = 2 \), there are 35 horizon plans with 210 total sequences to reach those 35 plans, or approximately \( 210/35 = 6 \) alternative sequences for each of the 35 plans.

The above numbers are, of course, only for a deterministic problem. For multiple state variable levels of number \( q \), the number of network simulations to be performed is increased significantly. Each alternative of the set \( n \) now becomes a set of alternative-state variable couples numbering \( n \cdot q \).

Given the former formal statement of the model and its assumptions, and the combinatorial nature of the decision tree we have just seen, there then remains the question of solution procedures for the general sequential decision model. Unfortunately, there has been very little published in this area. In fact, as Hadley [1] points out:

"There do not seem to be any references which treat in detail the procedures for solving the general sequential decision problem, although special types of problems receive attention in a number of works..."

The following sections will describe, first, a number of potentially feasible solution procedures for nominally reducing the space of alternatives by changing explicit exhaustive search to an implicit exhaustive search, and then the proposed procedure which ignores guaranteeing an exhaustive search, implicit or explicit, but concentrates on generating good efficient solutions to the time-staging problem.

136
3.3 Techniques for Solution (Alternative Search Procedures)

The potential solution techniques which could prove useful for the sequential decision model can be divided roughly into two major groups: direct and indirect search procedures. Direct procedures attempt to produce an optimum by a two-stage process. First, the benefit function is evaluated by calculating directly its value for different alternatives. Then, a new set of alternatives are chosen using a secondary criterion (which hopefully will improve the benefits obtained). Indirect methods attempt to choose the control variables (the alternatives) directly which will satisfy known conditions that produce optimality.

Probably the best example of indirect methods are the discrete programming methods of dynamic programming and branch and bound, or linear and non-linear programming techniques [3,4]. Because of the nature of the supply-demand equations and the simulation procedure for the transport problem, the discrete methods appear to be most appropriate of the indirect methods and will be the primary methods discussed here.

Direct methods can be divided into simultaneous and sequential search techniques. The former selects a series of alternatives to be tested a priori whereas the latter, iteratively, in a sequential fashion, attempts to improve on previous alternatives. The best example of the sequential direct approach is probably the elimination techniques of

---

These same categories exist for non-sequential problems. However, the number of techniques that are applicable to the sequential problem are far less than for the simpler non-sequential types. See Ochoa [2] for a good discussion of solution techniques for optimization problems.
successive approximation, Fibonacci, and golden section search, and the
direct climbing procedures of gradient techniques, pattern search or
creeping random search [5].

3.3.1 Indirect Mathematical Programming Search Procedures

Dynamic Programming

Possibly the most often applied and intuitively appealing tech-
nique of discrete indirect programming approaches, dynamic programming
offers several advantages for the general sequential decision problem that
most of the other solution procedures do not.¹ For example, application
of a dynamic programming formulation to the long-run investment problem
guarantees an optimal solution providing certain constraints are met.²

It also produces directly the optimal strategy from among the almost
infinite number of strategies that exist in an extremely efficient fashion.

The structure of a dynamic programming formulation is more a
philosophy than a particular search technique, however. Because of
this, there is usually more than one way to define a dynamic program
for the same problem. Different problems also generally require different
programs which precludes the use of any single general code such as those
that exist for linear programming. It is much more efficient to write a
separate program for each problem.

¹See Hadley [1,6], Nemhauser [7], Wagner [8], and Bellman and Dreyfus
[9] for an introduction to dynamic programming. See Howard and
Nemhauser [10], Mori [11], Bergendahl [12], Lai and Schaake [13],
Gulbransen [14], Funk and Tillman [15], and Taborga [16] for applica-
tions to long-run investment problems.

²We will discuss these constraints subsequently.
Based on Bellman's [17] principle of optimality, dynamic programming attempts to reduce an exhaustive, complete tree search by, first, noting that the path through the tree to a particular node k, will not affect the optimal path from node k on, and by then forming a series of recursive equations which can be solved to determine the optimal solution.

Its application to the long-run investment problem is relatively straightforward for both of the problems we will consider providing certain constraints are met. It is slightly more complicated when budget constraints are added, however, and including long-run supply-demand activity shift dependencies almost precludes dynamic programming as a solution technique. The objective of the dynamic programming formulation for the simpler problems of no long-run dependency is to find the optimal long-run investment sequence from among all possible sequence combinations. In terms of the project design case, or the network problem excluding budget constraints, we can define the deterministic investment problem in the following way:

---

1Bellman was the first to observe the generality of the principle of optimality and gave the procedure the name of dynamic programming. However, apparently the technique was used by Massé [18] as far back as 1944, and again by Arrow [19] in 1957. It is stated by Hadley [20] as follows: "...we cannot have an optimal value of the objective function for k stages unless for any \( x_k \) selected for stage k, the value of the objective function for the remaining k-1 stages is optimal, given the \( x_k \) selected for stage k."

2Or strategy, if returns are in some sense uncertain.
Let us define the set of permissible investment programs, $\xi_t$, for stage $t$ as

$$(\Lambda_t, \Lambda_{t-1}) \in \xi_t$$

which can be derived from the state transformation matrix, $T_A$.\footnote{We therefore suppress the subscript, $m$, denoting a particular plan, in the following formulations for ease of notation. Thus $\Lambda_{mt}$ and $\xi_{mt}$ are represented by $\Lambda_t$ and $\xi_t$, respectively.} Let the objective function be defined as [gross benefits - fixed cost] (discounted), summed over all periods $t$,\footnote{In this formulation, we are discounting both benefits and costs by a combined rate of discount, which balances a time preference discount rate with the marginal productivity of capital rate. A more general formulation would discount benefits by a social discount rate, $\rho_s$, and capital by the marginal productivity rate, $\rho_p$. The above equation would then be}

$$B = \sum_{t=1}^{N} \left[ g(\Lambda_t) - K(\xi_t) \right] (1+\rho)^{-t}.$$ \footnote{See Marglin [21] for a more detailed discussion of these discount factors.}

We can now form a forward-recursive dynamic programming function\footnote{This deterministic formulation is similar to Bergendahl's [22]. We are devoting some time to the dynamic programming formulation because of the similarity of the stochastic version and the model we use in the following sections.} (for the case of deterministic demands) which is defined for each stage in the process. That is, we focus on one stage at a time and
determine the optimal costs of operation (short-run costs) given the current network in place and the optimal longer-run costs of shifting from any previous network alternative to a new one.

The function is then defined, using the principle of optimality, as

\[
\text{The discounted net benefits for an optimal policy leading to network } \Lambda_t \text{ at time } t = \max_{\Lambda_t, \Lambda_{t-1}, t \in \mathcal{E}_t} \left[ \text{the benefits of optimal operation of plan } \Lambda_t \text{ at time } t \right] - \left[ \text{the investment cost of going from } \Lambda_{t-1} \text{ to } \Lambda_t \right] + \left[ \text{the discounted net benefits for an optimal policy leading to plan } \Lambda_{t-1} \text{ at time period } t-1 \right]
\]

or, in equation form: 1

\[
B_t(\Lambda_t) = \max_{(\Lambda_t, \Lambda_{t-1}) \in \mathcal{E}_t} \left[ (G(\Lambda_t) - K(\xi_t))(1+\rho)^{-t} + B_{t-1}(\Lambda_{t-1}) \right]
\]

s.t. \((\Lambda_t, \Lambda_{t-1}) \in \mathcal{E}_t\)

\[
B_0(\Lambda_0) = 0
\]

and

\[
B_{N+1} = \max_{\Lambda_N} \left[ B_N(\Lambda_N) + \Gamma_N(\Lambda_N) \right]
\]

1 Although this is forward-recursive, since the returns are deterministic, the function could have been defined as either forward- or backward-recursive. The optimal solution will be the same in either case, but one approach may be more efficient than another for some problems.
In the more general case where demands are stochastic, however, this formulation is invalid; we cannot provide a conditional optimal first period solution using the principle of optimality and a forward-recursive process. Instead we must proceed to the last stage of process and ask what is the optimal decision for stage $N$, given we are currently in state $m$, or in the transport case, the current alternative is $\Lambda_{m,N-1}$. We ask this question for all states $m$ (all alternatives $\Lambda_{m,N-1}$) and proceed recursively backwards down the tree.

The backwards-recursive formulation, for the case of an uncertain demand (and return) structure, is then defined as follows:

Let $\phi_t = \text{uncertain demand parameter set with probability}$

of occurrence $p_t(\phi_t)$.

$\bar{G}(\Lambda_t, \phi_t) = \text{the gross benefits of operation at stage } t \text{ with}$

$\Lambda_t \text{ in place when demand takes on a specific value}$

of $\phi_t$.

$\bar{B}_t(\Lambda_t) = \text{maximum expected benefits of an investment program,}$

$\Lambda_t$, for period $t$, when network in place at beginning

of $t$ is $\Lambda_t$.

$\hat{\xi}_t(\phi_{t-1}, \Lambda_t) = \text{optimal investment program for period } t \text{ given demand}$

in previous period was $\phi_{t-1}$ and network in place at

beginning of $t$ is $\Lambda_t$.

The (backward) recursive functions are then given as:

$$\bar{B}_t(\Lambda_t) = \max_{\Lambda_t, \Lambda_{t-1}, \xi_t} \left\{ [\bar{G}(\Lambda_t, \phi_t) p_t(\phi_t)] - K(\xi_t) (1+r)^{-t} \right.$$

$$+ \bar{B}_{t+1}(\Lambda_{t+1}) \right\} \text{ for } t = 1, \ldots, N-1$$

142
\[ s.t. \quad \Lambda_t, \Lambda_{t-1} \in \xi_t \]

\[ B_0(\Lambda_0) = 0 \]

and

\[ \overline{B}_N(\Lambda_N) = \max_{\Lambda_t, \Lambda_{t-1} \in \xi_t} \left\{ \left[ \sum \tilde{C}(\Lambda_t, \phi_N) \cdot p_N(\phi_N) \right] - K(\xi_t) \right\} (1+\rho)^{-N} \]

\[ + \Gamma_N(\Lambda_N) \}

The solution procedure begins by determining the expected net benefits \( \overline{B}_N(\Lambda_N) \) for stage \( N \) for each possible network, \( \Lambda_N \), in place at the beginning of stage \( N \). That is, we proceed to the last stage in the tree and average over the demand variable determining maximum expected benefits given the network in place is \( \Lambda_N \). We do this for each possible network \( \Lambda_N \) at stage \( N \). Simultaneously, we determine \( \hat{\xi}_N(\phi_{N-1}, \Lambda_N) \), the optimal investment program for period \( N \) given demand in the last period was \( \phi_{N-1} \), and the network in place at the beginning of \( N \) is \( \Lambda_N \). Continuing, the function \( \overline{B}_N(\Lambda_N) \) is used to determine \( \overline{B}_{N-2}(\Lambda_{N-1}) \) which in turn is used to determine \( \overline{B}_{N-3}(\Lambda_{N-2}) \) and so on until, at last, we determine \( B_1(\Lambda_1) \) using the existing, in-place network \( \Lambda_1 \). This then allows us to determine the best initial investment program \( \xi_1^* (\cdot, \Lambda_1) \).

An optimal strategy in terms of Section 2.3 is thus given by:

\[ S^* = \{ \xi_1^*, \hat{\xi}_2(\phi_1, \Lambda_2) \ldots \hat{\xi}_N(\phi_{N-1}, \Lambda_N) \} \]

where \( \xi_1^* \) is the optimal first period investment program. Each of the remaining terms are a set of programs, \( \xi_t(\cdot) \), conditional on the actual demand level, \( \phi_{t-1} \), in the previous period, and the network in place at the beginning of period \( t \).
Note that although this backwards approach has solved the problem of an uncertain demand structure, it still requires a number of extremely restrictive assumptions:

1. returns from each stage are separable, i.e., the return from any stage \( t \) does not depend on the path taken to reach that stage;

2. the end stage, \( N \), is finite and known;

3. state variables are discrete and independent from period to period (\( \phi_t \) does not depend on \( \phi_{t-1} \)).

Other than these assumptions, the solution has no other restrictions; the returns from each stage may be given by a linear flow model, by a simple equation or a complex simulation model. There are no other constraints such as monotonicity of the returns, linearity, etc.

It is also possible to reduce or relax a number of these constraints. For example, Devaney [23] in a recent thesis explored the concept of non-separable dynamic programming. In effect, he defines the optimal value function from the last point in the process (forward to \( N \)) where returns were separable. This solves the non-separability problem but can lead to horrendous sized dynamic programs.

There are problems of even broader implications, however, for the transport investment case. Dynamic programming, in order to guarantee an optimal solution, must search over all alternatives and all state variable levels. With any more than two or three state variables, and two or three stages, the amount of computation can become fairly burdensome.

Secondly, as we outlined in Chapter 1.0, in general, we cannot usually assume that the value of an end-state is independent of the path
of investments (and demands) used to reach it. In other words, if there are significant long-run dependencies between supply and demand, the assumption of separability no longer holds. (And non-separable dynamic programming, although an appealing approach and potentially applicable, is still in an infant stage of development.)

Finally, demands in one period are often not independent of the demands in previous periods, i.e., there is some correlation between \( \phi_t \) and \( \phi_{t-1}, \phi_{t-2}, \ldots \), and so on.

For problems with independent stage returns, a relatively small number of alternatives and state variables, the procedure is an efficient search procedure well adapted to the problems of investments under uncertainty. For the more general transport investment problem, however, dynamic programming will most likely be useful only as a partial search mechanism within a larger framework.

We have presented the dynamic programming formulation in some detail, however, because, although not completely and universally applicable, it does have some philosophically similar characteristics with our proposed procedure, to be presented in Section 3.4.

**Branch and Bound**

Potentially the most useful of the indirect methods, this procedure has seen little application to sequential decision problems and practically no application to realistic transport problems.\(^1\) The appli-

---

\(^1\)Ochoa [24], Ichbiah [25], and Bivins [26] have explored various branch and bound formulations and applied their algorithms to a variety of problems. Ochoa has shown that for the traditional knapsack problem, branch and bound is much more efficient than a dynamic programming solution procedure. Hershfordfer [27] has applied it to single period transport investment problems.
cations that do occur in the literature are generally in the area of single period and deterministic demand problems. Nothing has dealt with either stochastic or elastic demand functions, or multi-time period models, much less all three together, as we are attempting. We shall not attempt a formal algorithm here, but instead will roughly sketch out the basic concepts of branch and bound for a single period and then describe how it might be applied to the sequential investment problem in transportation.

Assume there are a series of projects, \( \ell_i \), \( i = 1...n \), each with a construction cost \( K_i \), and a given supply function. Demands are known, fixed and independent of the network in place. The total funds for construction are \( \zeta \), and the optimal set of projects for construction from the \( n \) potential additions are to be selected so as to minimize the total user cost over the network. Thus, we may define problem \( P \), after Ochoa and Silva [28], as:

\[
P = \text{Determine } \ell^0 \text{ and } \zeta^0 \text{ so as to minimize } \sum_{i=1}^{n} f(\ell_i) \]

Subject to

\[
\sum_{i=1}^{n} K_i \cdot \ell_i \leq \zeta
\]

\[
\ell_i = 0 \text{ or } 1
\]

The problem is similar to a great many capital budgeting formulations, except that the objective function is an implicit function of the flow volumes. That is, a dependency exists in the objective function since links are topologically connected and affect the flow volumes (and costs) for any given network. The objective function is evaluated by performing

\[^1\text{Alternatively, we could maximize net benefits such as consumer surplus minus operating costs.}\]
a traffic assignment of fixed demands to the existing network and some set of link additions of the complete link addition set, $2^n$. One solution technique, computationally infeasible for even moderate $n$, is to exhaustively explore all possible networks. A more efficient and intuitively appealing approach is to incorporate the procedures of discrete programming, either branch and bound or implicit enumeration (branch and backtrack).

The branch and bound algorithm for the solution of this problem is an iterative technique which produces a directed tree by successively branching from the best node to date and bounding the solution space. At each stage of the iteration, two new directed arcs and nodes are generated creating two more potential branching points. For each iteration, the set of branching nodes is divided into feasible and infeasible nodes, and the terminal node from which to branch at the next iteration is determined by the bounding procedure.

Without going into a formal description of the algorithm, we can briefly describe the solution technique as an iterative process of branching and bounding at terminal nodes (representing network alternatives) using network simulation procedures to produce the value of any node.

Roberts [29] has explored the application of Land and Doig's branch and bound algorithm to the single period network link addition problem for a developing country and has suggested a backward seeking search technique for the multi-time period problem. The first step in the multi-time period formulation is to proceed to stage $N$, the horizon, and solve for the best investment plan for period $N$ using the accumulated budget to
this point. Given this best investment plan, he then suggests that we iteratively step backwards in time, solving for each best investment for the preceding stage (again using the branch and bound algorithm) with a reduced budget, until we reach stage 1 and the existing network. The approach will produce a good investment staging sequence, although it cannot guarantee an optimal solution (even though demands are assumed to be independent of the staging path).\footnote{Notably it is similar in philosophy to the combined horizon-approximation, backwards-searching myopic rule of Section 4.2 that we propose, for the stochastic time-staged investment problem.}

For the details of a proposed, more complete branch and bound algorithm which will guarantee an optimal solution for the time-staged sequential investment problems, see Section 4.3 of Bivins [30]. Essentially, Bivins proposes to extend the single period formulation by defining decision variables, $l_i$, where

$$l_i = 0 \text{ implies that link } i \text{ is not built},$$

$$l_i = q \text{ implies that link } i \text{ is built in period } q,$$

and contingency constraints which force a link constructed in period $q$ to remain open for all other periods (disinvestment is not allowed). The problem now has $n \cdot N$ decision variables (although a set of $n$ decisions fully specifies a solution) and an enlarged number of branches to explore (there are $N$ branches possible from a node for each free link $i$). Demands again must be independent of the staging path to stage $N$.

Adding a stochastic structure to the formulation should not only increase the solution space (and quite dramatically), it should also
require a backward solution technique similar to dynamic programming. To the best of the author's knowledge, there have been no known multi-period stochastic branch and bound formulations to date, although with the current popularity and interest in these techniques we fully expect to see such formulations in the near future.

The solution method of branch and bound, however, still assumes that demands are independent of the staging path used to reach an end state and, in most cases, that demands are deterministic. For these reasons, we will not pursue this technique any further.

Other Indirect Techniques

There are a number of other indirect search techniques which can handle the sequential decision model in principle, such as chance-constrained programming [31], linear programming under uncertainty [32,33], stochastic programming with recourse [34], and so on, but they all require some fairly restrictive assumptions such as linear flow models, independence between stages, no more than two stages, or involve a tremendous number of equations which limits their usefulness for the problem we are concerned with.

3.3.2 Direct Search Procedures

In addition to the indirect methods of the previous section, there are also a number of simpler, direct procedures which can produce solutions to most transport investment problems but which cannot guarantee an optimal solution. For the most part, their application is rather straightforward for deterministic problems but considerably more complicated for problems of a stochastic nature. For example, for the problems
of the previous section, we need only select a number of investment program sequences \((\xi_1, \xi_2, \xi_3, \ldots \xi_N)\) and evaluate directly the net benefits and then compare to determine the best sequence. If we determine all sequences at once, then the process is labelled a simultaneous (direct) search procedure; if we sequentially select sequences, then it is a sequential search technique. This latter method may use a secondary criterion for selecting sequences such as a modified gradient criterion, or pattern search, and so on. Jacoby [35] used an iterative sequential sequence approach for power investments; there are many others found in the application areas of transportation, water resource, and other long-run investment problems [36,37].

For the case of a stochastic demand function, comparing selected sequences (whether simultaneously or sequentially selected) against random variables and averaging to produce an expected value can only produce the best investment sequence under uncertainty. With no option for revision, this is nothing more than the non-sequential investment problem under uncertainty as defined in Section 2.3.

As we described in the previous chapter, a number of solution procedures have arisen, mostly in the management field, which have attempted to circumvent this shortcoming which are classed as heuristic decision tree simulations. In most cases, there has been no single approach which has looked at producing strategies directly, other than in an unintentional heuristic fashion.

One way to develop strategies is to specify a priori conditional decisions which are then tested by sampling from the stochastic state.
variable space. For any sample state variable history, each of the conditional decision alternatives is then evaluated. This is repeated n times and the best strategy is then selected from among all of those tested. To the best of our knowledge, even this simple approach has not been applied, at least as reported in the literature.

Without delaying further on these direct techniques, we now turn to a final direct search procedure, the one we have chosen, and which we will apply (with modification) to the transport investment case.

3.4 A Decision-Tree Procedure for Transport Investments

Most, if not all of the procedures of the preceding section are limited in their ability to handle the application of the sequential decision framework to the transport time-staging problem. The reasons for this have been described elsewhere in this report a number of times. The purpose of this section is to describe a proposed computational procedure which can handle the most general form of the sequential decision model. The following sections describe the nature of this computational sequence and its application to transport investment problems. The basic structure is that of the extensive form decision tree calculations as used in statistical decision theory applications. This is described in detail in Raiffa and Schlaifer [38] and Raiffa [39]. The basic approach has been modified to reflect the nature of the transport investment problem. The use of a nonanalytic simulation model and a long-run supply-demand dependency, has required a special formulation and caused some additional implementation problems, which we will discuss in following sections.
3.4.1 Logic of the Tree-Tracing Procedure

As described in Section 3.3.1, deterministic decision trees are solvable by either forward or backward search procedures. Stochastic decision trees, on the other hand, can only be solved using a backward solution procedure: e.g., beginning with the last stage of the process, averaging over all state variables for each alternative, and then applying a decision rule for the optimal alternative at the last stage; the procedure then backs up one stage and repeats the process using the optimal value from the last stage as a terminal estimate for the current stage. This continues until the initial period is reached.¹

The reason for this backward search is fairly obvious. Demand, since it is a random variable, is unknown until it actually occurs. Therefore, we need not, and indeed should not, make a decision about investment at time t until the demand pattern history to that point is known. Of course, we do not know that history unless we actually wait for that point in time to occur. In order to make a decision at stage t, however, we can assume various possible "histories" and use the expected value principle to average over alternative histories to get an estimate of our future decisions.

More specifically, we assume a history (demand sequence) out to the last stage, and base our decisions for the last stage on the average

¹As we implied earlier, dynamic programming uses this backward induction principle by assuming separability between stages to reduce the number of branches to explore. In other words, it assumes the history preceding this point has no effect on future consequences other than the immediate preceding plan. It can therefore proceed to the last stage directly.
returns for this stage. These returns are averaged using the simulation results from a number of state variable levels, multiplied by the probability of those levels occurring. Repeating this for all stages and all possible histories allows us to then make an initial decision based on expected future values.

The actual sequence of these computations begins by first stepping forward in the process since the consequences of this terminal evaluation process depend on the history preceding this point. That is, long-run supply-demand interdependence (modelled in this study by a simplified activity shift model) produces dependencies between stages which requires us to first step out in a forward-stepping manner. This ensures that we know the demand history preceding this terminal averaging process, and have fully accounted for sequencing effects.

The actual computational procedure is described in flow chart form in Figure 3-5. It begins by choosing a first stage investment alternative, \( \Lambda_{m,t=1} \), and then a state variable demand level, \( \phi_{k,t=1} \). Given the investment pattern and demand level, it then evaluates the consequences of that couple by an explicit simulation run. The procedure then increments the time period by 1 and steps to the second stage to repeat the process, accumulating the results between stage 1 and 2 using a simple interpolation between stages. This stage-by-stage forward-stepping procedure is continued in this manner until the horizon or final period is reached. There, an investment alternative is chosen, and for each state variable demand level, a simulation is performed; the simulated consequences are evaluated and the net benefit is multiplied by its associated probability. This then
returns for this stage. These returns are averaged using the simulation results from a number of state variable levels, multiplied by the probability of those levels occurring. Repeating this for all stages and all possible histories allows us to then make an initial decision based on expected future values.

The actual sequence of these computations begins by first stepping forward in the process since the consequences of this terminal evaluation process depend on the history preceding this point. That is, long-run supply-demand interdependence (modelled in this study by a simplified activity shift model) produces dependencies between stages which requires us to first step out in a forward-stepping manner. This ensures that we know the demand history preceding this terminal averaging process, and have fully accounted for sequencing effects.

The actual computational procedure is described in flow chart form in Figure 3-5. It begins by choosing a first stage investment alternative, \( \Lambda_{m,t=1} \), and then a state variable demand level, \( \phi_{k,t=1} \). Given the investment pattern and demand level, it then evaluates the consequences of that couple by an explicit simulation run. The procedure then increments the time period by 1 and steps to the second stage to repeat the process, accumulating the results between stage 1 and 2 using a simple interpolation between stages. This stage-by-stage forward-stepping procedure is continued in this manner until the horizon or final period is reached. There, an investment alternative is chosen, and for each state variable demand level, a simulation is performed; the simulated consequences are evaluated and the net benefit is multiplied by its associated probability. This then
Define: Initial Network Configuration
   Alternative $\Lambda$
   Demand and Supply Functions
Initialize: Stage $i = 1$

For each stage $i$, Set demand level $\phi_{k=1}$

Consider alternative $\Lambda_j$ at stage $i$

Consider demand level, $\phi_k$ at stage $i$

Compute return for alternative $\Lambda_j$
at stage $i$ for demand level $k$

Compute discounted net benefits
   Perform activity shift

Transport Simulation Model

Increment stage $i = i + 1$  No

Last Stage? (i = N) Yes

Last demand level at this stage? No

Choose new demand level $k = k + 1$

Expected Value of Best $\Lambda_{ji}$, $E(\Lambda^*_{ji}) + Terminal Value$
   for previous $\Lambda_j, \phi_k$ at last stage

Last alternative at this stage? No

Choose new alternative, $j = j + 1$

Yes

Retract one stage $i = i - 1$

Back at stage $i = 0$ Yes

EXIT

Figure 3-5
Flow Chart of the General Tree-Tracing Algorithm

154
gives an expected final period value for the current sequence and final period alternative.

We repeat the procedure at the final period for each alternative, then compare alternatives, saving the best alternative for that stage. After we have exhausted the alternatives at this final stage, and chosen the best of these for this sequence, we then retrace our steps in the tree, backing up one stage and selecting a second state variable level for the alternative at stage N-1. Another simulation is performed and we again step out to stage N, repeating the averaging-out and folding-back procedure described above. This process continues, averaging out over state variable levels at a stage, then folding-back the tree to retrace its steps, choosing a new branch to explore, and continuing forward again out to the end-state. Given enough computation time and a reasonably small number of alternatives, the procedure will eventually arrive back at the initial state with the most desirable first period decision, and a series of subsequent decisions, conditional on the alternate histories that can occur.

3.4.2 A Formal Statement of the Algorithm

With that relatively brief description of the procedure, let us now define the steps of the algorithm, depicted rather approximately in Figure 3-5, in more detail.

Step 0.0 Simulation calibration. For t<1, simulate with a single known state variable level and existing network structure. Begin accumulation procedure using discount factors and interpolation procedures. Predict the demand shift for
period $t=1$. Specify permissible alternative sequences, state variable classes, temporary horizon $N$, and number of states. Set stage $t=1$, state variable class $k=1$, alternative index $m=1$.

**Step 1.0** Initialize: Set "best action" and "best value" of this stage equal to minus infinity. Select next most current alternative $\Lambda_{m,t}$ for stage $t$. Check if permissible transformation $\Lambda_{j,t-1} \rightarrow \Lambda_{m,t}$ by calling subroutine PERMIS which checks state transformation matrix, $T_{\Lambda}$. If permissible, proceed to Step 2.0. If not, proceed to Step 4.0 to see if analysis finished, or to choose a new alternative.

**Step 2.0** Select state variable class $k$, for uncertain parameter, $\phi_{kt}$. Simulate consequences by calling transport simulation model (MODEL). Check if at horizon $N$. If yes, proceed to Step 3.0. If no, continue.

2.1 Update period $t = t+1$; return to Step 1.0 to continue search over next stage.

**Step 3.0** Check if all state variable classes $\phi_{kt}$ at state $t$ have been simulated. If not, proceed to Step 2.0 with a new state variable level. Otherwise, continue.

3.1 Compute current best estimate of expected value of alternative $\Lambda_{m,t}$ at stage $t = KK$ using terminally evaluated future consequences, UBEST($N$), and interpolated stage returns STACU($KK$).
3.2 Compare expected return from alternative \( \Lambda_{mt} \) at stage \( t \) with best previous alternative: save the best alternative's index and its expected net benefits. Proceed to Step 4.0.

Step 4.0 Check if any additional alternatives \( \Lambda_{mt} \) to be simulated at stage \( t \). If none, check if back at stage 0 and analysis is finished. If \( t=0 \), proceed to Step 5.0. If \( t\geq1 \) fold back one stage, \( t = t-1 \), and proceed to Step 3.0. If there are some \( \Lambda_{mt} \) to be evaluated and compared, select a new alternative index \( m \) and proceed to Step 1.0.

Step 5.0 Output best initial investment pattern and conditional staged strategy.

It will be useful for later sections to explore the exact form of some of these procedures in more detail, such as the discounting and interpolation procedures, and the detailed probability calculations.

The Discounting Procedure

The procedure for discounting is a relatively straightforward process. At each stage \( t = KK \), when a net benefit, \( PVU(KK) \), is predicted based on the simulation of equilibrium for each \( (\Lambda_{mt}, \phi_{kt}) \) couple, the normal present value discounting procedure is applied to produce a discounted net present value, \( PVD(KK,I) \), for a range of interest rates, or

\[
PVD(KK,I) = PVU(KK) \cdot (1+I)^{-KK}
\]

for \( I = .05, .10, .15 \).
Interpolation Between Stages

Given the discounted values for two contiguous stages, t and t+1, we then approximate the full time stream of values by a linear interpolation scheme, shown in Figure 3-6. That is, the stage accumulated returns, STACU, are determined by summing the area for stage KK, again for different discount rates, I, as

\[
\text{STACU}(I, KK) = [\text{PVD}(I, KK-1) + \frac{\text{PVD}(I, KK) - \text{PVD}(I, KK-1)}{2}] \cdot \text{STAGE}(KK)
\]

where STAGE(KK) is the stage length.

This is then accumulated for all stages to produce the accumulated net present value, ACCPV(I, KK), up to stage KK by

\[
\text{ACCPV}(I, KK) = \text{ACCPV}(I, KK-1) + \text{STACU}(I, KK)
\]

The Detailed Expected Value Computations

The purpose of presenting the previous two computational techniques is as background for the detailed expected value calculations, which are unique to investment planning models. At any point in the tree, the value of all future expected discounted returns from stage KK+1 on are stored in UBEST(KK+1), shown in Figure 3-7. This is determined by using the net discounted value PVD(I, KK+1) at stage KK+1 and interpolating, either by assuming constant returns to \( \infty \), or using the interpolation procedure and PVD(I, KK+2).

The expected return for alternative, \( \Lambda_m^{t+1} \), the network under consideration for investment during period t to t+1 (it will be in place at stage t+1) is then computed for stage KK (KK to KK+1) as PROD(KK):

\[
\text{PROD}(KK) = \text{POST}(KK) [\text{STACU}(KK) + \text{UBEST}(KK+1)]
\]

158
Figure 3-6
Interpolating between Stages and Accumulating Net Present Values

159
Figure 3-7
The Detailed Expected Value Calculations
160
summed over all state variable levels, where $\text{POST}(KK) =$ the posterior probability of the current state variable $k$, and the other variables are as previously defined.

In other words, the expected net present value for stage $KK-1$ is stored in $\text{UBEST}(KK)$ and is computed as:

$$\text{UBEST}(KK) = \max \left[ \sum_{m, KK} \{ \text{STACU}(KK) + \text{UBEST}(KK+1) \} p_k(\phi_k, KK) \right]$$

\[ \Lambda_{m, KK} \phi_k, KK \]

3.4.3 A Set of Heuristic Approximations

One of the unfortunate consequences of this tree-tracing logic is that it does not take many stages, state variable levels or alternatives before the tree search soon gets out of hand. This is certainly true when, for each alternative plan, $\Lambda_{m,t}$, and state variable level, $\phi_{k,t}$, simulation of supply-demand equilibrium is required, as described in Section 2.2. Each point in the tree of this type takes approximately 10 seconds of IBM 360/67 execution time for a network of 34 links. With a reasonable number of alternatives and state variable levels, this could soon stretch into hours of CPU time. For a real world problem, the analysis procedure of Section 3.4.1 will experience difficulty in just reaching an end-state, let alone begin averaging-out and folding-back.

Two approaches to this problem are (1) to structure the search in a more efficient manner using programming methods which make use of conditions which produce optimality and implicitly prune large numbers of alternatives, or (2) to introduce approximations to the complete tree-search procedure.
Programming methods attack the combinatorial problems of multi-period investment problems but require fairly restrictive assumptions (as discussed previously), as in integer linear programming, or dynamic programming. They also still require extensive computational time and cannot guarantee even a good feasible solution.

Therefore, we have chosen the second approach in order that we may both insure that solutions are produced, and control the amount of computational effort expended. That is, we apply the decision analysis framework but incorporate approximating procedures to reduce the space of solutions. The design of these procedures is described in detail in the following chapter. First, we briefly explore the efficiency of three search techniques.

3.5 A Comparative Analysis of Search Efficiencies

A comparison of three of the search techniques most useful for the time-staging investment problem should demonstrate their relative efficiencies and provide some insight into the limitations of their constraints, implied or otherwise. The three procedures chosen for comparison are direct enumeration, decision analysis, and dynamic programming. As an example, consider the tree diagram of Figure 3-8, which could represent a multi-time period (deterministic) investment problem. The tree has three stages with a number of different states at each stage. (Note that the stages are numbered in reverse order for the dynamic programming recursion.) Stage 1 has five "states," stage 2 has four, and stage 3 has three. The tree also has a number of overlapping paths. State 3 at stage 4, for

162
Figure 3-8
An Example Decision Tree for Comparing Search Efficiencies
example, can be reached by four distinct paths, each of which is part of another path. Each arc or stage path has a unique value associated with it, which represents a distance. This would represent a simulated stage return at the end of each arc in the transport problem.

Let \( N \) = maximum number of stages = 3

\[ K_n = \text{number of state variables at each stage} \]

\[ J_n = \text{decision variables at stage} \ n \]

There are

\[ K_N^N \prod_{i=1}^{N} J_n \]

feasible solutions or distinct paths for this problem, or \( 1(3 \times 2 \times 2) = 12 \) feasible solutions.

By direct enumeration of all 12 paths (the tree is shown in Figure 3-9), the search involves taking each path separately, evaluating the consequence of this path and performing pair-wise comparisons to determine the best path from among the 12. For each path, we require, therefore, the addition of \( N \) numbers, one for each stage up to the maximum stage, \( N \).

By adding 2 numbers at a time, there are

\( (N-1)K_N \prod_{i=1}^{N} J_n \) additions

or

\( (3-1) \ 1(3 \times 2 \times 2) = 24 \) additions.

Given each total path value, to determine the best path requires

\( (K_N \prod_{i=1}^{N} J_n) - 1 \) comparisons

or

\( 3(3 \times 2 \times 2) - 1 = 11 \) comparisons.
Figure 3-9
The Decision Tree for Comparing Decision Analysis and Exhaustive Enumeration

165
Therefore, the total number of operations for direct enumeration is equal to the number of additions plus the number of comparisons or

\[ 24 + 11 = 35 \text{ operations}. \]

Using the same logic, decision analysis reduces the number of operations because it uses the averaging-out and folding-back procedure described in Section 3.4.1, which reduces the number of additions and comparisons. Climbing the tree as described in Section 3.4.1, we observe that decision analysis requires only 19 additions and 10 comparisons for a total of 28 operations.

Recursion analysis (dynamic programming) goes one step further and observes that a number of the end states in Figure 3-9 are common to a number of paths. For stage 1 through \( N \) there are

\[ \sum_{n=1}^{N} K_n \text{n additions} \]

or

\[ (4x2) + (3x2) + (1x3) = 17 \text{ additions}. \]

In addition, for each state variable, there are \( J_n - 1 \) comparisons up to stage \( N \). At stage \( N \) there are \( K_N - 1 \) comparisons, or the total comparisons equal

\[
[ \sum_{n=1}^{N} K_n (J_n - 1)] + K_N - 1 \\
= (4x1) + (3x1) + (1x2) + (1-1) = 9
\]

for a total number of operations equal to \( 17 + 9 = 26 \) operations.

Note the significant reduction in number of operations for both decision analysis and dynamic programming over direct enumeration. This will be further amplified if uncertainty is allowed (i.e., at the end of 166
each arc there are a number of possible outcomes instead of a single value). The difference between recursion and decision analysis is minor for this example. For larger trees, this difference will become more pronounced. However, note that dynamic programming requires an assumption which neither of the other two require: that stage returns are separable - the return from one stage does not depend on the path taken up to this point.

If, however, returns from one stage do depend on the path taken as we will show in Section 5.2 (as will be the case of most investments), then normal separable dynamic programming is inappropriate, and we must turn to either non-separable dynamic programming or decision analysis.

3.6 Summary and Conclusions

The purpose of this chapter has been to present a more mathematical formulation of the transport investment problem, especially for the case of uncertainty of demand, and to explore in detail a number of potential solution techniques for this problem; most notably, we first explored formulations incorporating the discrete optimization techniques of branch and bound and dynamic programming. We concluded that these procedures will be useful for the general investment problem primarily as partial search techniques, and that very little has been done to date with either approach for the investment problem under uncertainty. More importantly, neither technique handles the proposed characteristics of the transport
problem outlined in Chapters 1.0 and 2.0 (especially, when there is long-run dependence between supply and demand).

We further extended a forward-recursive deterministic dynamic programming formulation to a backward-recursive dynamic programming version of a restricted problem (i.e. no long-run dependencies), because of its similarity to the procedure we propose.

We then presented the tree-tracing technique we propose to use as a solution method for the general sequential decision model applied to the transport investment problem and compared the efficiencies of three decision tree procedures: exhaustive search, dynamic programming and the proposed procedure of decision analysis. Dynamic programming is the most efficient but is much more restrictive in its assumptions. Exhaustive search is clearly out of the question because of the number of alternatives it implies. Decision analysis is much improved in terms of the amount of computation over exhaustive search and is much less restrictive in the assumptions it requires for solution over dynamic programming.

We noted, however, that there still will be computational problems with any more than a few stages and a reasonable number of alternatives and state variables. This provides the motivation for the proposed heuristic procedure of this thesis.

We now turn to a detailed presentation of the heuristics we propose in Chapter 4.0.
Chapter 4

A PROPOSED HEURISTIC SOLUTION PROCEDURE FOR THE
SEQUENTIAL DECISION FRAMEWORK
Chapter 4

A PROPOSED HEURISTIC SOLUTION PROCEDURE FOR THE SEQUENTIAL DECISION FRAMEWORK

The purpose of this chapter is to describe first, the approximating techniques necessary to make application of the sequential decision model to the transportation investment problem feasible, and second, how these techniques have been implemented in a series of computer programs designed to test the effects of these procedures. These procedures are then applied to a typical single link capacity expansion problem in Chapter 5.0.

We begin the chapter with a brief summary of the analysis problem and search philosophy followed by the proposed heuristic decision analysis procedure. Section 4.3 then presents a formal statement of the algorithm. Section 4.4 turns to some additional complications, aside from those that arise from applying a sequential investment framework to transportation planning, such as selecting an initial horizon and how to circumvent distorting edge effects of that selection. And finally, we present a summary of the approach, in Section 4.5.

4.1 Description of the Approach

In a previous chapter, we described how the sequential decision framework overcomes one of the most obvious deficiencies of existing transport planning approaches, that of ignoring both uncertainty and the sequential nature of investment implementation. The last chapter then discussed possible alternative solution techniques for this sequential
framework, including the proposed procedure of a multistage stochastic decision tree in extensive form. This procedure was chosen because of the unique characteristics of the general transportation investment problem (i.e., a dependence exists between stages, between supply and demand, returns are to be specified by a non-analytic simulation model, and so on, as outlined in Chapters 1.0 and 2.0) and the ability of the proposed procedure to handle these characteristics.

The objective of the stochastic model based on the sequential decision framework is to produce the optimal first stage investment decision by exploring conditionally all combinations of alternatives over a number of probabilistic demand levels. The procedure for comparing strategies is the averaging-out and folding-back technique which uses the expected value principle and a maximizing rule to reduce the timestream of values to a discounted single point estimate for each initial investment. Application of this procedure produces what we have called the global solution.

In principle, although we are primarily concerned only with the initial investment, the objective of the sequential decision framework is to explore each alternative's implications as far into the future as possible. However, the combinatorial effects of exploring stage after stage of alternatives, conditional on previous state variable levels, can soon get out of hand, since the number of combinations increases more than linearly as the number of stages increase.

Therefore, although the extensive form gets around most of the former problems (such as dependence between stages, etc.), its solution
presupposes a substantial amount of computer time and, more importantly, a finite decision tree. Unfortunately, for any investment problem with a reasonable number of alternatives, state variable levels, and time periods, the tree is essentially an infinite one.

Practically speaking, then, we need to be concerned with a means to reduce the number of combinations to search; to achieve a balance between computing each alternative's complete future implications, and the costs of doing so. This can be handled in a number of ways. For example, since we are actually only concerned with the initial decision, we need not explore all possible sequences or all combinations of alternatives for all demand levels. In many instances, we only need a reasonably good estimate of the future consequences associated with any investment alternative in order to be able to discriminate between alternatives at the initial stage.

We define the problem $p_k$ to be the sub-problem of determining the optimal investment strategy and its expected value from any node $k$ on the tree (see Figure 4-1). The problem $p_k$ can in many cases be approximated by an equivalent smaller problem $p'_k$, which gives us a reasonably good estimate of the optimal decision at this stage and its expected net present value. The computational problem then becomes one of determining when a portion of the decision tree can be approximated and what value of expected utility should be placed on that point as an estimate of the characteristics of the sub-tree.

There are two aspects to this problem: first, we need a procedure to tell us when a particular sub-problem can be approximated by a
Figure 4-1
Approximating Parts of the Decision Tree by a Sub-Problem $p'_k$
reduced sub-tree; these we define as pruning rules. Second, given we are
going to approximate the problem \( p_k \), we need some mechanism to estimate
the expected utility of this sub-tree without actually performing the cal-
culations; these functions we define as terminal evaluation functions.

We have implemented these functions as a series of specific sub-
routines to be used with the overall decision tree program, DECISN. The
logic of the computations essentially follows the structure of the general
sequential decision analysis model of Section 2.3, the hierarchically-
structured sequential decision model of Manheim,\(^1\) and the flow chart out-
lined in Chapter 3.0. The specific details of cutting the investment de-
cision tree—the specific pruning rules and terminal functions we have
employed—are unique to the transport problem, however. The following
sections describe these elements in more detail, how they are to be used,
and when they are most appropriate in any given problem.

4.2 The Proposed Heuristic Search Procedure

4.2.1 Determining the Cut-off Point Using Explicit Pruning Rules

There are a great many different pruning rules that can be used
to reduce the number of calculations, although any particular rule also
depends in part on the terminal evaluation function to be employed. In
effect, a pruning rule decides when enough is known about a particular

\(^1\) We are indebted to Professor Manheim for his insight into this problem
and his initial work in this area. See Manheim [1] for a discussion of
pruning rules and terminal functions for a hierarchically-structured
prescriptive model applied to the highway route location problem.
sequence of investment-state variable couples in order that a terminal function can be called upon to estimate the future consequence of problem $p_k$ from this point on.

Each rule, therefore, must be aware of the accuracy with which a terminal function can estimate the consequences of the removed sub-tree. This accuracy will be difficult to predict a priori. Within any given problem, we expect the knowledge of the effectiveness of a particular set of pruning rules to increase, as well as from problem to problem.¹

The following is a list of pruning rules that have been implemented and are currently operational in DECISN:²

1. Probability Limit, PMIN(t): this procedure cuts off the extensive tree search when either
   
i. the stage probability of a state variable $p_k(\phi_{kt})$ is so low that its "expected" contribution to the expected net benefits at this stage is either negligible or will have no effect on the decision at this point in the tree, i.e., the tree is pruned and a terminal evaluation function is called when $p_k(\phi_{kt}) < PMIN(t)$, or

¹We have not incorporated any automatic function for learning about or updating these rules in the current set of programs, but this is a fairly straightforward extension.

²These rules are similar in principle to those employed by Manheim [1]. The primary difference between the two approaches is in the exact form of terminal evaluation function for the transport problem, once a pruning rule indicates computations should stop at a particular stage.
ii. the path product probability of any sequence contributes negligibly to the expected value of future investments at stage $t$. If we define $H_t = [\phi_{kt}] = [\phi_{k1}, \phi_{k2}, \ldots, \phi_{kt}]$ as the history of observed results to stage $t$ and if the probability of the state variable, $\phi_{kt}$, at stage $t$ is $p_k(\phi_{kt} \mid H_{t-1})$, then the total probability of the history to stage $t$ is given by

$$T = p(\phi_{k1}) p(\phi_{k2} \mid \phi_{k1}) p(\phi_{k3} \mid \phi_{k2}, \phi_{k1}) \ldots p(\phi_{kt} \mid H_{t-1})$$

$$= p(\phi_{k1}) \prod_{i=2}^{t} p_k(\phi_{ki} \mid H_{i-1}).$$

A terminal function is called when this path product probability is less than the minimum probability set for this stage---when $T < \text{PMIN}(t)$. Thus, if the probability at any stage, whether the stage or the accumulated path product probability, is less than some criterion, $\text{PMIN}(t)$, the normal tree tracing procedure stops and we proceed to call the terminal evaluation functions.

2. The current horizon ($N$):\footnote{In Section 4.4, we discuss the initial choice of the horizon, $N$, further.} this sets a rather arbitrary and tentative limit to the period of interest. Once this specific period is reached (by either proceeding normally in the manner of the sequential step-by-step "climbing" of the tree, or by a previous application of a terminal
function, \( \Gamma_k(\Lambda_{mt}, \phi_{kt}) \), we stop the computations, invoke a simple form of the set of terminal evaluation functions, \( \eta_k(\Lambda_{mt}, \phi_{kt}) \), (see section 4.2.2), and proceed by folding back to the previous stage.

The above pruning rules and the related terminal functions are shown schematically in Figure 4-2. The initial solid line branches at stage 1 represent alternative mutually exclusive networks or projects \( \Lambda_{mt} \). The dashed lines are events or uncertain state variables, \( \phi_{kt} \) (in the case of the transport investment problem they represent an uncertain demand parameter set), which produce a set of consequences for each alternative-state variable couple by means of a complex simulation procedure.

If, after simulation at stage \( t \), the set of pruning rules indicate that a full tree search is unnecessary beyond this point, a terminal evaluation function is invoked, represented by the cross-hatched circles (points A, B, and C, for example), which estimates the consequences of the tree from this point on by some approximating procedure. If pruning is not indicated the tree-climbing procedure continues, selecting alternative-state variable couples and performing a simulation until either a constraint is violated or the tentative horizon, \( N \), is reached. In this diagram, we have shown terminal functions being invoked at various points in the tree for three pruning criteria (stage probability at point A, path probability at B, and the horizon constraint at C).

In summary, the proposed procedure iteratively steps out in time, selecting action-state variable couples for each stage, simulating and then evaluating the partial consequences by multiplying the probability
Terminal functions: $\Gamma_k(\Lambda_{m,t},\phi_{kt})$  \hspace{1cm}  $\Gamma_j(\Lambda_{m+1,t+1},\phi_{i,t+1})$  \hspace{1cm}  $\eta_m(\Lambda_{m,N},\phi_{i,N})$

Stages:  Stage 1  \hspace{1cm}  Stage 2  \hspace{1cm}  Stage 3

Figure 4-2
Pruning Rules and Related Terminal Evaluation Functions
178
of that state variable by the result, then repeating until either

1. the horizon is reached, or

2. a pruning rule determines that the main probability
criterion is violated for
   (a) individual stage probabilities
   (b) stage probability products.

Once the decision has been made to stop searching down this branch, the future must then be evaluated by an appropriate terminal evaluation function.

4.2.2 Adjusting for Variable Pseudo-Horizon Cut-Offs Using Terminal Evaluation Functions

Once a pruning rule indicates that the search should stop exploring a particular branch, we then call for a terminal evaluation function, $\Gamma_k(\Lambda,\phi)$, to produce an estimate of the pruned portion of the tree. Similar in principle to Jacoby's terminal correction factors [2], these functions attempt to correct for a less than complete tree search of future actions and outcomes by employing simplified approximating functions.\footnote{The analogy is only partially valid, since Jacoby is concerned with a fixed-horizon, deterministic model and with the distorting edge effects of this form of model on the investment decisions. His functions attempt to adjust for the N-th period asset structure passed onto the future.} Their purpose is to produce an estimate of expected utility placed on this node, as if the full decision tree computation were performed, but without actually completing the computations. These functions then act as surrogates for the full decision tree computation.
These functions are shown in Figure 4-3 for any stage less than an arbitrary horizon limit, N, for this particular run. In this diagram, a terminal function is represented as a single cross-hatched circle (for example, point A in the diagram). In some cases, it may be a simple mathematical function which acts as the estimating function. For example, at point A in the figure, we have reached the pruning rule horizon limit, N. This point can be expected to fall at a fairly distant point in the tree (see Section 4.5) and we use a simple form of the set of the terminal functions, labelled \( \eta_k(\Lambda_{mN}, \phi_{kN}) \). In the present version of DECISION, we have the following options for \( \eta_k(\Lambda_{mN}, \phi_{kN}) \):

1. Assume no future from this horizon point on.

2. Assume constant returns from N to \( \infty \) using the estimate of returns at stage N.

3. Assume a simple percentage growth of returns, either linear or exponential, equal to the average growth over the last k periods.

More generally, however, terminal evaluation functions will be called before reaching N. In this case, the function will require a more complex set of procedures to produce a good estimate of the accumulated future expected utility. These functions may, in fact actually be yet another decision tree computation. For example, at point B in the diagram we indicate a terminal function, \( \Gamma_j(\Lambda_{n,t}, \phi_{j,t}) \), has been called. The larger lower circle encompassing the tree beginning at point B, represents a second level decision tree calculation. This calculation also has a number of options. In other words, we may have a simplified decision tree
These functions are shown in Figure 4-3 for any stage less than an arbitrary horizon limit, N, for this particular run. In this diagram, a terminal function is represented as a single cross-hatched circle (for example, point A in the diagram). In some cases, it may be a simple mathematical function which acts as the estimating function. For example, at point A in the figure, we have reached the pruning rule horizon limit, N. This point can be expected to fall at a fairly distant point in the tree (see Section 4.5) and we use a simple form of the set of the terminal functions, labelled \( \eta_k(\Lambda_{mN}, \phi_{kN}) \). In the present version of DECISION, we have the following options for \( \eta_k(\Lambda_{mN}, \phi_{kN}) \):

1. assume no future from this horizon point on.
2. assume constant returns from \( N \) to \( \infty \) using the estimate of returns at stage \( N \).
3. assume a simple percentage growth of returns, either linear or exponential, equal to the average growth over the last \( k \) periods.

More generally, however, terminal evaluation functions will be called before reaching \( N \). In this case, the function will require a more complex set of procedures to produce a good estimate of the accumulated future expected utility. These functions may, in fact actually be yet another decision tree computation. For example, at point B in the diagram we indicate a terminal function, \( \Gamma_j(\Lambda_{n,t}, \phi_{j,t}) \), has been called. The larger lower circle encompassing the tree beginning at point B, represents a second level decision tree calculation. This calculation also has a number of options. In other words, we may have a simplified decision tree
Figure 4-3
Embedded Terminal Evaluation Functions
181
embedded in a larger tree. This terminal function (the smaller tree) may also have its own pruning rules and terminal functions represented in the diagram by point D. In theory, the procedure could have terminal functions embedded within terminal functions embedded within other terminal functions and so on, ad infinitum. In practice, we do not expect more than three or four levels of these functions. The approaches which have been utilized in DECISION involve a number of procedures which are presented in detail in the following sections. The properties of both these functions, and of the pruning rules of the last section, will then be explored in detail in the following chapter.

4.2.2.1 A Collapsed Stage Space

The first form of terminal evaluation function acts on the number of periods in the model and is termed a stage aggregation function. This function collapses a multi-period model into a more aggregate form and interpolates between stages in order to estimate the results of intermediate stages. For example, at point B in Figure 4-3, although the full tree search would involve searching over the three periods remaining from point B to the horizon (each of five years in length), the stage aggregation terminal function, STAGG, reduces the size of the tree by collapsing the three periods to a smaller number of periods; perhaps a two-period model of five, and 10-year stage lengths, or even a one-period model in some cases. All discounting procedures, growth factors, etc., are automatically aggregated and adjusted for any stage-length increase once pruning is indicated.

182
For relatively small incremental changes in demand, the use of stage aggregation will produce fairly good estimates of the expected net benefits for any point in the tree to a given horizon. As shown in Table 3-6 of Section 3.2.7, for 20 potential link investments with a maximum of one link addition per period, reducing the periods from 10 to two, for example, reduces the number of time-staged investment sequences from greater than $5 \times 10^{11}$ to 380.

4.2.2.2 The Myopic Rule

The second major terminal function is the myopic search rule. At a pruning point, the expected utility of future decisions can be estimated by an investment sequence provided by performing a myopic stage-by-stage selection procedure. This procedure looks ahead only one stage at a time, comparing all alternatives of the current stage, and then branching to the next stage from the previous stage's best alternative.\(^1\)

By definition, a myopic rule provides the solution to a sub-problem. Instead of sequentially climbing the tree from period to period with a new alternative-state variable couple at each stage in the normal averaging-out and folding-back technique of the global search procedure, the myopic rule ignores all future periods and selects the best action based only on the current stage. This best action can also be determined in two ways: either deterministically, or using the expected value principle if demands are stochastic.

\(^1\)A variant on this approach is to conduct a myopic k-level search. That is, use the myopic rule over a total of k stages.
With that best action, the procedure then uses the mean state variable as the node to proceed from for both the deterministic and stochastic cases. All other tree climbing from the dominated actions at stage \( k \) is ignored. The effect on the number of computations is evident from Figure 4-4.

The consequences of such a rule can be dramatic for both number of computations and for results, depending on where and when the rule is used. Its effect is a stage-by-stage optimization of the objective function, which can, by definition of the problem's time related dependencies, lead to costly investment decisions if used at the wrong time.

As we noted earlier, however, for some portions of the tree fairly distant from the initial period, a complete exploration of the tree will be unnecessary. All that is required is a relatively good estimate of the utility from that point on. There are a number of investment strategies that can produce a good estimate. In some cases, the myopic rule can produce a good strategy with little computational effort.

We also note that in some instances, the myopic rule can even produce the optimal strategy. For example, if discount factors are high, so initial benefits are weighted proportionately more than future benefits, or if economic lives are very short, a decision produced by the myopic rule can be identical to one produced by the global rule.

In general, however, because of its inherent error as a sub-optimizing procedure, we expect this procedure to be useful primarily only in the extreme branches of the tree which contribute a relatively minor amount to the objective function. It may also be extremely useful when
Figure 4-4
Comparison of Myopic and Global Rules
coupled with another procedure such as the stage aggregation technique of the previous section.

4.2.2.3 A Reduced State Variable Approach

One of the more common procedures in the past for handling uncertainty has been to assume a continuous distribution over the uncertain variable (such as the normal or gamma distribution) and to sample randomly from this distribution to produce an expected value. In practice, however, unless integration or a complete monte carlo random sampling procedure is to take place (which for the transport investment case is obviously out of the question), we must make a discrete approximation to the continuous case. We use such an approach in DECISN by employing a discrete probability distribution \( p_k(\phi_{kt}) \) with \( k \) elements or classes (see Figure 4-5(a)).

As an approximation to the complete probabilistic decision tree computation, therefore, another obvious approach to terminally evaluate a point without an exhaustive exploration of the tree is to further reduce the number of classes of the state variable distribution from a fairly extensive number to a reduced set; for example, from ten or 20 classes to two or three as shown in Figure 4-5(b).

The extreme case of this approach is to treat the problem as a deterministic one from the pruning point on— that is, perform the normal simulation procedures out to the current horizon, \( N \), but using a distribution with the number of classes equal to one. The most appropriate value for this single value is either the element with maximum probability or a value representing the mean state variable level.
Figure 4-5
Discrete Approximations to the State Variable Distribution
For single stage problems and simple cost and benefit structures, it is relatively simple to determine this single value, termed the certainty equivalent\(^1\) which will yield the identical result as when the full distribution is used. For multi-period models and for non-linear functions, the best value is not so obvious nor as easy to derive.

We define the use of a single state variable search as a deterministic search; a single state variable sequence is defined as a spine. Again, as in the myopic rule, this procedure can lead to erroneous conclusions since we are approximating a continuous probability function by only selected values of the variables. However, in some cases, either a deterministic problem which uses only the mean state variable level, or a low and high state variable spine coupled with the mean state variable spine can give fairly good results as estimating procedures in place of the full tree exploration. It is also possible to couple this procedure with either the myopic procedure or that of stage aggregation. The resulting difference in numbers of calculations is shown schematically in Figure 4–6 for a selected mean state variable spine. If not coupled with other procedures, it is then still classed as a global rule, but with a reduced state variable space.

4.2.2.4 Approximating the Alternative Space at a Terminal Point

The final class of approximations centers on the selection procedure itself at any stage. To this point, the search procedure has assumed

\(^1\) See Raiffa and Schlaifer [3], page 177, for a discussion of the use of certainty equivalents.
Stage 1

Figure 4-6
Selected Mean State Variable Spines
that all alternatives of the set are exhaustively compared. For the project design problem or small network problems, this may be an acceptable procedure. For a problem with any more than a few potential alternatives, an exhaustive selection procedure is clearly out of the question. As we showed earlier, for example, for n=11, the number of possible plans at a stage is $2^n = 2044$.\footnote{Recall from the previous chapter that the actual number at a stage may be less than this, since the set of plans possible at any stage may be constrained by the plan at the previous stage, i.e., some investment programs will be physically impossible.}

Thus, we need a mechanism for reducing this number. There are a number of procedures which are possible. The simplest, but probably least effective, is a random selection of a number of plans from the full set of $2^n$. A more useful and directed approach would involve iterative direct search procedures, which select plans based on past performance, i.e., links which appear to be the most congested. In other words, we use the results of the simulation results at previous stages to suggest which links to try at the current stage.

As a variant of this approach, we might initially reduce the alternative plans at the current stage using the stage aggregation function as a look-ahead heuristic in what we have called the horizon approximation technique (see Figure 4-7). This approach initially steps to some future period, possibly the initially selected horizon, N, with all periods in between aggregated into one stage using the stage aggregation function. A partial search, using the same iterative direct search technique we just described, can be used to produce a number of good plans at the horizon,
Figure 4-7
The Horizon-Approximation Sub-Problem
much less than the full $2^n$ in number. Restricting the alternatives at a previous stage to be a subset of these horizon plans before beginning a search there will also further reduce the number. This would then operate as a single-period model or as a horizon approximation procedure similar to Roberts [4], Salazar and Sen [5], and Consad [6]. In the more general case, we might extend the horizon approximation to include some aggregated stages in between in order to better capture the effects of time-sequencing.

The danger in all these techniques, of course, is that we cannot be sure we have not eliminated the best alternative, since we are ignoring the sequencing effects. We have not incorporated alternative selection procedures at a stage in the current version of DECISION although this also would be a straightforward extension. We can, however, use the stage aggregation function as a horizon approximation scheme and heuristically select the best $k$ alternatives of the set $2^n$ for evaluation at a previous stage. Clearly, there are any number of other procedures which would be useful in reducing the alternative selection procedure. We will discuss this further for the network context in Chapter 7.0.

4.3 The Adjusted Tree-Climbing Algorithm

The proposed procedures can now be stated as coded within the programming system, DECISION, as adjustments to the formal algorithm of Section 3.4.2. Figure 4-8 is a more detailed flow chart for the overall tree-tracing logic described in Section 3.4.1 with these adjustment procedures included.
Figure 4-8
Flow Chart of DECISION's Tree-Tracing Algorithm

193
3.0 Compute current best estimate of expected value of alternative $A_{mt}$ at stage $t$ using terminally evaluated future consequences, UBEST(t+1) and interpolated stage returns

3.1 Check if all state variables at this stage have been tested

3.2 Compute Net Expected Future Return for Alternative $A_{mt}$

3.3 Compare Expected Return from this alternative $A_{mt}$ at stage $t$ with best previous alternative; Save best alternative

4.0 Are there any alternatives left to be simulated and compared at stage $t$?

4.1 Check if back at stage 1? Is analysis finished?

4.2 Select new alternative index $m = m+1$

5.0 Output best initial investment pattern and strategy

Figure 4-8 (Cont'd)
194
Essentially, the algorithm performs the tree-tracing logic by sequentially selecting alternatives and state variables as long as a constraint is not violated. Once that does occur, however, a set of routines is called which performs a sub-tree search. With that estimate complete, the procedure again proceeds to climb the tree and eventually begins the averaging-out and folding-back procedure described earlier, until it eventually arrives back at the base of the tree. A formal statement of the algorithm represented in this flow chart follows:

Step 0.0 Initialize data and select an initial search strategy, specify permissible alternative sequences, state variable classes, temporary horizon N, and number of stages.

0.1 Simulation calibration. For $t = -1, 0$, simulate with a single known state variable value and existing network structure. Begin the stage accumulation procedure using discount factors and interpolation procedure. Predict the demand shift for period $t = 1$.

0.2 Set stage $t = 1$, state variable class $k = 1$, alternative index $m = 1$.

Step 1.0 Set best action and best value of this stage equal to dummy alternative and large negative number, respectively.

1.1 Select next most current plan alternative, $\Lambda_{mt}$ for stage $t$.

1.2 Check if permissible transformation $\Lambda_{j,t-1} \rightarrow \Lambda_{mt}$ by calling subroutine PERMIS which checks state transformation matrix, $T_{\Lambda}$.
If permissible, proceed to Step 2.0. If not permissible, proceed to Step 4.0 to see if analysis finished, or to select a new alternative index.

**Step 2.0** Determine current state variable class, \( k \), for uncertain parameter value \( \phi_{kt} \).

1. Simulate consequences by calling transport simulation model, MODEL, and compute returns.

2.2 Check if at tentative horizon, \( N \). If yes, call horizon terminal evaluation function, \( \eta_k(\Lambda_mN, \phi_{kN}) \); then proceed to Step 3.0. Otherwise, proceed to Step 2.3.

2.3 If not last stage, check if tree should be pruned at this point; call PRUNRL. If tree to be pruned, proceed to Step 2.4. If not, proceed to 2.5.

2.4 Determine type of terminal evaluation function, \( \Gamma_k(\Lambda_{mt} \phi_{kt}) \), to be used at this point: stage aggregation, myopic search, horizon approximation, maximum probability spine, etc. For example,

(i) if stage aggregation indicated, call STAGG, aggregate periods from \( t \) to horizon \( N \) into single period, mark current stage by MARKK, proceed to Step 2.5.

(ii) if myopic search strategy indicated, determine whether single state variable or stochastic myopic search. If deterministic, aggregate classes of state variable, \( \phi_{kt} \), for stage \( t \) into single maximum probability spine, call routine MYOPDN as terminal evaluation routine, proceed to Step 3.0.
2.5 Update the period $t = t+1$; return to Step 1.0 to continue search over next stage, whether normal or stage aggregated tree.

Step 3.0 Compute current best estimate of expected value of alternative $A_{m_t}$ at stage $t$ using terminally evaluated estimate of future consequences, $UBEST(t+1)$, and interpolated stage returns.

3.1 Check if all classes of the state variable, $\phi_{kt}$, at stage $t$ have been simulated.
If not, proceed to Step 2.0 with a new state variable level. Otherwise continue.

3.2 Compare expected return from alternative $A_{m_t}$ at stage $t$ with best previous alternative; save the best alternative's index and its expected net benefits; proceed to Step 4.0.

Step 4.0 Check if any additional alternatives $A_{mt}$ are to be simulated at stage $t$. If none, go to Step 4.1. Otherwise, go to Step 4.2.

4.1 Check if back at stage 1 and analysis is finished: if $t = 1$, proceed to Step 5.0. If $t > 1$, fold back one stage, $t = t-1$, proceed to Step 3.0.

4.2 Select a new alternative index $m$ and proceed to Step 1.1.

Step 5.0 Output best initial investment pattern and strategy.
Output search strategy and pruning rule-terminal evaluation functions used in this run. Proceed to Step 0.0, select new search strategy.
2.5 Update the period \( t = t+1 \); return to Step 1.0 to continue search over next stage, whether normal or stage aggregated tree.

Step 3.0 Compute current best estimate of expected value of alternative \( \Lambda_{mt} \) at stage \( t \) using terminally evaluated estimate of future consequences, UBEST(\( t+1 \)), and interpolated stage returns.

3.1 Check if all classes of the state variable, \( \phi_{kt} \), at stage \( t \) have been simulated.

If not, proceed to Step 2.0 with a new state variable level. Otherwise continue.

3.2 Compare expected return from alternative \( \Lambda_{mt} \) at stage \( t \) with best previous alternative; save the best alternative's index and its expected net benefits; proceed to Step 4.0.

Step 4.0 Check if any additional alternatives \( \Lambda_{mt} \) are to be simulated at stage \( t \). If none, go to Step 4.1. Otherwise, go to Step 4.2.

4.1 Check if back at stage 1 and analysis is finished: if \( t = 1 \), proceed to Step 5.0. If \( t > 1 \), fold back one stage, \( t = t-1 \), proceed to Step 3.0.

4.2 Select a new alternative index \( m \) and proceed to Step 1.1.

Step 5.0 Output best initial investment pattern and strategy.

Output search strategy and pruning rule-terminal evaluation functions used in this run. Proceed to Step 0.0, select new search strategy.
4.4 Application of the Procedure to the Transport Investment Case

4.4.1 Uniqueness of Transportation: Irreversibility and Longevity

Most of the problems caused by the combinatorial effects of multiple time periods and uncertainty can be traced to the fact that investments, especially those of a transport nature, have fairly long economic lives and hence fairly significant future consequences. Once in place, it is difficult to imagine the process reversing itself. That is, transport is the type of investment in which disinvestment is difficult to accomplish in most instances. Walters [7] states the case succinctly when he says:

"Once we have built a road it is sunk in a particular location for a long time. Furthermore, and this is where it differs from the majority of investment models, the irreversibility occurs in increasing the capacity of the roads.... In one sense the irreversibility argument is nothing more than the old adage 'bygones are for ever bygones'."

Essentially, this longevity and irreversibility is what ties the individual stage decisions together, from period to period. It is this same longevity coupled with the growth in demands which requires us to consider future decisions and their impact on initial ones, and requires a global averaging-out and folding-back procedure.

If, as Cole has assumed [8], the periods or stages can be limited to coincide with the economic lives of our facilities, we may reduce the combinatorial problem by using a myopic stage-by-stage optimization and be assured that this will produce the global optimum (even for the case of uncertain future outcomes). But this is difficult to do normally and almost begs the question of sequential decision making. In other words,
the problem, if defined in this manner, is coincident with the static uncertainty, single period problem. The initial decision is independent of future decisions but a staging pattern based on the expected value for each stage can be chosen once and for all.

4.4.2 Selection of a Horizon

The fact that there are considerable impacts of future decisions and events on immediate decisions is the major reason for considering investments over multiple time-periods. However, future costs and benefits are weighed proportionately less, the further into the future they occur. Therefore, beyond a certain point in time, the effects of the future can be assumed to be negligible. This point is usually difficult to find in practice and is often beyond what is normally chosen as a fixed horizon for modelling purposes.

One of the major arguments put forth for selecting a short horizon is that normal discounting procedures of the present value technique will favor current benefits over future benefits in such a manner, that at a 50-year horizon, and a rate of 5%, the value of a dollar is reduced to approximately $0.08. Therefore, there is no need to consider planning with excessively lengthy horizons. The value of benefits will be approaching zero at a horizon much beyond 30 years.

Unfortunately, this approach fails to recognize at least two factors. First, we are not so much interested in the absolute value of an investment at 30 years as we are in the value at that time relative to other investments, and its effects on immediate decisions. And secondly, considering an increasing discount factor alone fails to recognize that
benefits are going to be growing, simply by increases in population alone. Therefore, if population, and the propensity to travel both increase, as has historically been the case, these effects would tend to negate the reducing effect of discounting—or at least offset it enough so that the absolute (discounted) benefits are quite substantial for a longer period of time.

Some of the other commonly noted factors which are influential in the selection of a fixed horizon period\(^1\) are:

1. the economic life of a project,
2. the limitation of our predictive capabilities in predicting future consequences of an investment, i.e., the validity of our forecasts,
3. the effects of larger studies and other sectors of the economy with different horizons than the current transport study.

In fact, the further one proceeds into the future, the more uncertain the demand forecasts, and the more we need to know how a system will operate under a variety of conditions. It is exactly in this kind of situation when investment should occur in more flexible and adaptable systems. Assuming either certainty models or no future beyond a fixed point can and will bias the evaluation of alternative patterns of investment.

In summary, a number of factors affect the modelling problem of horizon selection. The ultimate goal is to select a cut-off point such

\(^1\)See Jacoby [9], Kendrik [10], Wohl [11], Massé [12], and Weingartner [13] for further discussions of the problems of horizon selection.
that extending the calculation beyond this point has no appreciable effect on the initial investment decision. The limitation of this approach is in trying to decide \textit{a priori} where this point occurs. The further into the future we select it, the more we have increased our chances for determining the best initial period investment. On the other hand, the greater the horizon, the greater the simulation time and computing costs. The use of a long horizon is more accurate, but can soon limit the usefulness of the model. Therefore, the usual practice has been to select some period as a fixed horizon beyond which the future is ignored (generally, in transportation, this has been 20 to 30 years into the future). This limits the number of calculations and puts alternative comparisons on a comparable scale but unfortunately, because of long economic lives, it can lead to severe distortions in the investment pattern in some instances.

4.4.3 \textit{Adjusting Terminal Conditions for Distorting Edge Effects}

The conclusions one draws from the arguments of the previous section point to the selection of a fixed horizon model, if for no other reason, for computational efficiency alone. However, the use of such an approach can, in some instances, cause severe distortions in the final investment pattern, as we have just stated. A cut-off point at the horizon without some form of adjustment will tend to favor less capital intensive technologies in the later periods, since any investment made at this stage will have its capital charged to this period and yet be unable to count any long-term benefits from this investment. Because of investment interdependence, this in turn will affect the investment choices in earlier periods.
There are a number of approaches useful for adjusting for a fixed horizon. Kendrick [14] has used one approach in his study of programming investments in the process industries. In effect, he converts the capital cost to an equivalent annual uniform series value and the payments are then cut off at the end of the period of the model. The system, therefore, need only pay for that portion used within the fixed horizon.

Another approach is that of Chakravarty and Eckaus [15], which forces the terminal capacity to be sufficient to maintain post-terminal rates of growth of consumption, or exports, or other appropriate variables. Russel [16] has used another tack, simpler but probably less accurate for determining initial investments. He simply fixes the investment for the last period to equal the existing capacity deficit, recognizing earlier investments and deficits. This approach really ignores the problem of distortion and defines a boundary condition which limits the number of alternative investment paths, but which may introduce biases of its own.

The approach of adjusting for terminal conditions is also similar in many respects to the salvage value technique of engineering economics used by Bergendahl [17] in his network investment problem and many others. Other approaches (Manne [18], Muhich [19]) resolve the issue by assuming an infinite horizon and use mathematically tractable approaches to solve for the scale and timing of single projects.

An alternative to an extended horizon approach is one we mentioned earlier, suggested by H. Jacoby [20] in a thesis on the analysis of investment in electric power. He suggests, and develops, a set of terminal correction factors, similar to our terminal evaluation functions. They
are designed to adjust the "returns" of any investment plan if it deviates from a base case of capacity at the horizon. Thus, the distorting edge effect of penalizing large investments in later periods is avoided. However, this is essentially a post-simulation correction procedure with strategies of investment having been exogenously specified and consequences predicted and then adjusted. There is no driving force for selecting good alternatives directly, although conceivably some technique of this sort could be incorporated as an automatic procedure.  

We have chosen to attack the problem of horizon selection using a number of techniques, most notably the equivalent uniform payment series approach of Kendrick (originally attributable to Stephen Marglin; see Kendrick [21]), coupled with the terminal evaluation functions described in Section 4.2.2. Capital costs are first converted to equivalent annual costs, $\gamma_{mt}$, such that the sum of the annual discounted costs for each year of the facility's life, $L_m$, will just equal the original total investment of that facility occurring in year $t_m$, $K(\xi_{mt})$, or

$$K(\xi_{mt}) = \sum_{t=t_m}^{L_m+t_m} \gamma_{mt} (1+\rho)^{-t}$$

In effect, $\gamma_{mt}$ is determined by applying the capital recovery factor, CRF, to the total investment sum, $K(\xi_{mt})$, or

---

1 In essence, the search procedure is an iterative direct search procedure whereby the user specifies alternative sequences which are then tested, terminally corrected, and compared to previous sequences.
\[ \gamma_{mt} = \left[ CRF_{\rho, L_m} \right] [K(\epsilon_{mt})] \]

\[ = \left[ \frac{(1+\rho)^n}{(1+\rho)^n - 1} \right] [K(\epsilon_{mt})] \]

For each year simulated, the amount \( \gamma_{mt} \) is then discounted back to the base year, \( t_0 \), by the normal discounting procedures described in Section 3.2.4. The effect of this procedure is to charge each facility with only a portion of its total capital if a portion of the facility's life falls outside the current horizon. If we continued with a lump sum charge of capital, those projects in later periods would be charged with a large incremental cost and have no chance to enjoy the future benefits that will occur over that facility's life. As we may be incorporating a different effective horizon for each branch of the tree when we employ a pruning rule and terminal function, the approach becomes especially appealing as a technique to circumvent distorting edge effects.

4.5 Summary of the Search Procedure

One of the primary purposes of this thesis has been to develop a flexible search procedure for producing time-staging strategies for the transport investment problem, recognizing explicitly both the use of a non-analytic simulation model to predict the consequences of these investments, and the uncertainty of future demand projections and its effect on the investment pattern. We have done so using the general sequential decision model as the basic model framework.

Unfortunately, the application of this general sequential decision model to the transport case usually results in severe computational
problems because of the extreme dimensions of the solution space.
Therefore, procedures for reducing these dimensions have been
developed from two viewpoints. First (and this has been the primary
emphasis of this chapter), we have recognized the extreme size of
the tree and the need for sequential search procedures and have
therefore developed a set of techniques called pruning rules and
terminal evaluation functions for reducing the space of solutions.
And second, we have been concerned with developing good heuristic
investment strategies a priori, which are extremely flexible and
able to adapt to unforeseen changes in the environment. We will
discuss this further in the following chapter.

In summary, the proposed procedure iteratively steps out
a stage at a time. At each stage, it first selects an alternative-
state variable couple and then simulates the flow pattern expected
from such a couple. It then evaluates the partial consequences by
multiplying the probability of that state variable by the result,
then repeats the procedure until either

1. the horizon is reached, or

2. a pruning rule determines that the mean probability
criterion is violated for

   (a) individual stage probabilities

   (b) stage probability products

Once pruning is indicated, the sub-tree emanating from this point
is estimated by calling a terminal evaluation functions. For points
sufficiently far out, this is represented by a special form of
terminal function, \( \eta_k(\Lambda_{mN}, \Phi_{kN}) \). For points closer in, a slightly different and more complex procedure is required, labelled \( \Gamma_k(\Lambda_{mt}, \Phi_{kt}) \) which can involve a variety of techniques, either singly or in combination.

In order to keep the computational effort within the scope of the research effort, we have selected three or four of these pruning rules and terminal evaluation functions to program in order that we may test their properties and evaluate their usefulness for the transport investment problem. Using these programs, we have performed a comparative analysis in an example problem for the project design case which is presented next in Chapter 5.0.
Chapter 5

COMPUTATIONAL EXPERIENCE WITH THE PROPOSED SEQUENTIAL MODEL
Chapter 5

COMPUTATIONAL EXPERIENCE WITH THE PROPOSED SEQUENTIAL MODEL

The presentation of the general sequential decision model and its application to the transport investment case has been completed at this point. This presentation, although designed for transportation investment problems in general, including both project design and program selection or network problems, has assumed that the underlying stochastic structure can be represented by a known probability distribution. Relaxing this assumption, and the development of a model for adaptively revising the investment pattern and our estimate of the underlying uncertainty, will be explored in Chapter 6.0.

The purpose of the current chapter will be first, to investigate the issues of investment planning under uncertainty for the case of a single link or project design problem with a known probability distribution, and second, to explore the properties of a selected number of pruning rules and terminal evaluation functions as applied to this same problem. By initially relaxing a number of constraints, we hope to present our computational experience with the proposed procedure and to learn something about the properties of the model that will prove useful in the more complex problems. As we have already shown in Chapter 2.0, even the supposedly simple problem of

208
single link investment is, in fact (for the transport case) a network problem, as soon as any additional new investment is considered.

The chapter first presents the example data in terms of costs, supply and demand functions, activity shift, and benefit measures in Section 5.2. We then present our computational experience with the model of Chapter 3.0 for a relatively "small" problem taken from the literature. Thus, we can initially ignore the approximating procedures of Chapter 4.0. The purpose of this section is to (a) perform a comparative analysis of the results of alternative planning philosophies using deterministic, non-sequential stochastic and sequential stochastic model formulations and (b) observe the effects on the investment pattern for staging and economies of scale tradeoffs, without introducing the additional complications of computational approximations.

Section 5.4 then applies the pruning rules and terminal evaluation functions to the same example in order to explore the properties of these procedures and to assess the usefulness of applying a general sequential decision framework to the long-run investment planning problem. We conclude the chapter with a discussion of computation times, and a summary of the results and the conclusions we have reached for the project design case.

5.1 Introduction: General Description of the Problem

The specific example chosen to demonstrate the proposed procedure is similar to the problem of project design found in the field of highway investment planning [1,2]. However, the example
we have chosen to present is somewhat different from most typical project design problems. We have extended the traditional assumptions and define the problem characteristics as follows:

(1) The supply of, and demand for, transport service for any given short run situation is assumed to be highly interdependent.

(2) Shifts in the demand for transportation, represented by shifts in the underlying socio-economic variables of the demand function, are also assumed to depend heavily on the facility sequence implemented.

(3) The solution of supply-demand equilibrium (the determination of the flow pattern) is given by a complex, nonanalytic and iterative simulation procedure.

(4) The demand for transport service is subject to some degree of uncertainty. That is, the estimate of demand can either be assumed to be unknown to some degree, or subject to an underlying random structure.

In the example presented here, we assume the project is a portion of highway located in a newly populated corridor and is primarily used for commuting purposes. We also assume the primary volumes are between the two end points A and B and have no alternate path (see Figure 5-1).

The investment alternatives proposed to alleviate existing congestion include improving the existing facility, location 0, an at-grade low-volume facility, to a higher capacity, limited-access freeway of varying design standards, or to build a new facility in one or more locations (H or G). In addition, each location H and G may be constructed to two capacity levels or scales; a 2 or 4 lane facility. Each location also has the option of partial or staged
<table>
<thead>
<tr>
<th>Location</th>
<th>Alternative</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>Existing 4 lane undivided at-grade roadway</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Improve alternative 1 to divided, limited access facility, level 1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Improve alternative 1 to divided, limited access facility, level 2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Improve alternative 1 to divided, limited access facility, level 3</td>
</tr>
<tr>
<td>H</td>
<td>5</td>
<td>Construction of an additional 2 lane facility on new r.o.w.</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Construction of alternative 5 with additionally buying r.o.w. for 4 lanes</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Add an additional 2 lanes to make alternative 5 a 4 lane, grade separated facility</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>Add an additional 2 lanes to alternative 6</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>Immediate construction of 4 lanes, grade separated on new r.o.w.</td>
</tr>
<tr>
<td>G</td>
<td>10</td>
<td>(Same as alternative 5 but in Location G)</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>(&quot; &quot; &quot; &quot; 6 &quot; &quot; &quot; &quot;)</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>(&quot; &quot; &quot; &quot; 7 &quot; &quot; &quot; &quot;)</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>(&quot; &quot; &quot; &quot; 8 &quot; &quot; &quot; &quot;)</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>(&quot; &quot; &quot; &quot; 9 &quot; &quot; &quot; &quot;)</td>
</tr>
</tbody>
</table>

Figure 5-1
A Project Design Example
211
construction, i.e. to build 4 lanes of the 4 lane facility now, but only paving two, building only 2 lanes now but buying right-of-way for four and so on. In all, there are 14 distinct projects to consider in this initial example as shown in Figure 5-1.¹

We also assume that alternatives H and G are mutually exclusive i.e. either one or the other may be constructed to any scale but not both. This is typical of situations where residential development has not reached either location H or G, and deciding on either is relatively simple and inexpensive at the present time, but once that location is decided, development will either preclude the other location or make it extremely costly.

The problem then is to decide on the optimal time-staged investment program for the corridor considering both present and future supply and demand.

5.2 Structure of the Model

5.2.1 Supply Functions

The supply curves for each of these 14 projects are shown in Figure 5-2. They represent the unit price or impedance that a user will experience in making the trip from A to B over each of the different facilities at various volume levels. This price includes all perceived user costs including operating costs, congestion, time

¹We have limited this initial example to 14 in order to explore the model's properties in a fairly simple and uncluttered problem. Obviously, there can be many more alternatives by simply increasing the number of locations and/or the number of partial constructions.
Legend:  
A Existing Alternative 1  
B Existing Alternative 1, Improved to Level 1  
C Existing Alternative 1, Improved to Level 2  
D Existing Alternative 1, Improved to Level 3  
E Alternatives 5, 6; 10, 11 - 2 lanes  
F Alternatives 7, 8, 9; 12, 13, 14 - 4 lanes

Figure 5-2  
The Link Supply Functions

213
and discomfort costs, as described in Section 2.2. For example, alternative 1 is the existing 4 lane, undivided, at-grade facility and has fairly high operating costs for all levels of flow. The form of this function for alternative 1 and each of the other alternatives is as follows:\textsuperscript{1}

1. Existing 4 lane unimproved roadway: Alternative 1

\[ p_k = 7.0 + \frac{350.0}{31.0 - 0.013q_k} \text{ cents/trip} \]

where \( p_k \) = total price of trip on existing roadway per mile (including taxes)

\( q_k \) = hourly volume per direction

2. New 2 lane facility: Alternatives 5, 6, 10, 11

\[ p_2 = 9.0 + \frac{350.0}{46.6 - 0.0226q_2} \text{ cents/trip} \]

3. New 4 lane facility: Alternatives 7, 8, 9, 12, 13, 14

\[ p_4 = 9.0 + \frac{350.0}{46.6 - 0.0113q_4} \text{ cents/trip} \]

Alternatives 2, 3 and 4 are simply improved alternative 1 supply functions. They represent various levels of upgrading the design characteristics of alternative 1.

\textsuperscript{1}These functions are similar to those presented in Section 2.3 and to those developed by Wohl and Martin [3].
All of the 14 alternatives can be represented by one or more of these six functions. In other words, even though alternatives 5 and 10 are in different locations, we are able to model their supply characteristics with the same curve since they are both 2 lane facilities of equal length. (They will, however, have quite different capital and maintenance costs which makes them unique.) If either location was different in some way (length, alignment or profile, etc.) it would be a simple matter to include additional supply functions to account for these differences.

5.2.2 Aggregate Supply: Network Effects for the Project Design Case

Although locations 0 and H are mutually exclusive, alternative locations 0 and G or 0 and H are not. Project 1, the existing facility will continue to operate even if a new facility in locations H or G is constructed.¹ (If project 2 is selected, the existing facility merely operates at an improved price level for all volumes.) We therefore define the aggregate supply each user will face in order to determine equilibrium between supply and demand. The aggregate supply for the existing facility (1) and a new 2 lane facility (5) is shown schematically in Figure 5-3. It represents the impedance or price levels that exist when more than one facility is available

¹This assumes that we have eliminated the alternative of abandoning the existing facility. This is a possible alternative, but in most cases, it is not generally considered. We will ignore it for purposes of this example.
Figure 5-3
Forming the Aggregate Supply Function for Two Parallel Links
and it is constructed by summing horizontally each of the supply curves of the two alternative projects.

In terms of the total number of mutually exclusive alternatives, and the plan investment matrix of Chapter 3.0, the 14 distinct projects of our example now results in 44 mutually exclusive (aggregate supply) plans, \( \pi^m \), to be tested (shown in Table 5-1).

Given these aggregate supplies, the assignment of flows is handled in the following manner: First, the equilibrium flow volume is determined for the given demand and aggregate supply conditions as shown in Figure 5-3. In order to determine how total flow will distribute itself, we must then specify a flow distribution mechanism. In this case, we use Wardrop's second principle [4], which assumes that flows distribute themselves between competing capacity such that prices (or costs) on the two facilities are equivalent. With this assumption and the equilibrium price computed, the volumes on each link contained in the aggregate supply are given by \( q_A \) and \( q_E \) in Figure 5-3.

5.2.3 The Demand Function

A common approach in most project analysis studies is to assume that demand is either exogenously specified or given by simple growth extrapolation, and that it is inelastic or independent with respect to the supply offered. Since our primary concern will ultimately be with the network problem, we assume that the demand for travel between zones A and B is similar to formulations used in the literature for network
<table>
<thead>
<tr>
<th>Project No.</th>
<th>Plan No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>&quot;O&quot;</td>
<td>&quot;C&quot;</td>
<td>&quot;H&quot;</td>
<td>&quot;O&quot;</td>
<td>&quot;C&quot;</td>
<td>&quot;H&quot;</td>
<td>&quot;O&quot;</td>
<td>&quot;C&quot;</td>
<td>&quot;H&quot;</td>
<td>&quot;O&quot;</td>
<td>&quot;C&quot;</td>
<td>&quot;H&quot;</td>
<td>&quot;O&quot;</td>
<td>&quot;C&quot;</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<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>2</td>
<td>1</td>
<td></td>
<td></td>
<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>3</td>
<td>1</td>
<td></td>
<td></td>
<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>4</td>
<td>1</td>
<td></td>
<td></td>
<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>5</td>
<td>1</td>
<td></td>
<td></td>
<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>6</td>
<td>1</td>
<td></td>
<td></td>
<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>7</td>
<td>1</td>
<td></td>
<td></td>
<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>8</td>
<td>1</td>
<td></td>
<td></td>
<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>9</td>
<td>1</td>
<td></td>
<td></td>
<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>10</td>
<td>1</td>
<td></td>
<td></td>
<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>11</td>
<td>1</td>
<td></td>
<td></td>
<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>12</td>
<td>1</td>
<td></td>
<td></td>
<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>13</td>
<td>1</td>
<td></td>
<td></td>
<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>14</td>
<td>1</td>
<td></td>
<td></td>
<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>15</td>
<td>1</td>
<td></td>
<td></td>
<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>16</td>
<td>1</td>
<td></td>
<td></td>
<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>17</td>
<td>1</td>
<td></td>
<td></td>
<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>18</td>
<td>1</td>
<td></td>
<td></td>
<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>19</td>
<td>1</td>
<td></td>
<td></td>
<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>20</td>
<td>1</td>
<td></td>
<td></td>
<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>21</td>
<td>1</td>
<td></td>
<td></td>
<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>22</td>
<td>1</td>
<td></td>
<td></td>
<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>23</td>
<td>1</td>
<td></td>
<td></td>
<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>24</td>
<td>1</td>
<td></td>
<td></td>
<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>25</td>
<td>1</td>
<td></td>
<td></td>
<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>26</td>
<td>1</td>
<td></td>
<td></td>
<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>27</td>
<td>1</td>
<td></td>
<td></td>
<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>28</td>
<td>1</td>
<td></td>
<td></td>
<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>29</td>
<td>1</td>
<td></td>
<td></td>
<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>30</td>
<td>1</td>
<td></td>
<td></td>
<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>31</td>
<td>1</td>
<td></td>
<td></td>
<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>32</td>
<td>1</td>
<td></td>
<td></td>
<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>33</td>
<td>1</td>
<td></td>
<td></td>
<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>34</td>
<td>1</td>
<td></td>
<td></td>
<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>35</td>
<td>1</td>
<td></td>
<td></td>
<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>36</td>
<td>1</td>
<td></td>
<td></td>
<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>37</td>
<td>1</td>
<td></td>
<td></td>
<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>38</td>
<td>1</td>
<td></td>
<td></td>
<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>39</td>
<td>1</td>
<td></td>
<td></td>
<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>40</td>
<td>1</td>
<td></td>
<td></td>
<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>41</td>
<td>1</td>
<td></td>
<td></td>
<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>42</td>
<td>1</td>
<td></td>
<td></td>
<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>43</td>
<td>1</td>
<td></td>
<td></td>
<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>44</td>
<td>1</td>
<td></td>
<td></td>
<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>
</tbody>
</table>
problems and discussed in Section 2.3. The only difference is that
the technology characteristics or level-of-service variables have been
collapsed into a single "price" variable using a linear weighting
function. The demand function for the example problem is then given
by the following formulation:

$$V_{ij} = \alpha_0 (P_{ij})^{\alpha_1} (Y_i)^{\alpha_2} (Y_j)^{\alpha_3} (c_{ij})^{-\beta}$$

where

$$c_{ij} = c_{ij} + t_{ij} \cdot v + w_{ij} + x_{ij}$$

anc  \( V_{ij} = \) interzonal demand between zones i and j

\( P_{ij} = \) population of zones i,j

\( Y_i = \) average per capita income of zone i

\( c_{ij} = \) perceived price of travel or impedance between i and j

\( o_{ij} = \) out-of-pocket vehicle operating costs

\( t_{ij} = \) travel time between i and j

\( v = \) value of time

\( w_{ij} = \) expected accident costs for the trip between i and j

\( x_{ij} = \) vehicle tax for the trip between i and j

\( \alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta = \) model parameters

Population and income data for the demand model are given in
Table 5-2 for the two zones.
Table 5-2
Population and Income Data

<table>
<thead>
<tr>
<th>Zone</th>
<th>Population 1965</th>
<th>Population 1970</th>
<th>Income Average Per Capita ($/Year) 1965</th>
<th>Income Average Per Capita ($/Year) 1970</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>500,000</td>
<td>510,000</td>
<td>2220</td>
<td>2414</td>
</tr>
<tr>
<td>B</td>
<td>100,000</td>
<td>120,000</td>
<td>1870</td>
<td>2032</td>
</tr>
</tbody>
</table>

The regional growth of population and income over time is assumed to be specified from a larger interregional study. Income growth is taken as 3 1/2% per year, and population growth as 2% or 3% per year in the following examples.

Fitting the model to existing population, income, and volume data results in the parameter set, \( \phi = [\alpha, \beta] \), which reproduces the volumes for the first two calibration periods. We assume the demand model parameters remain constant for future predictions. Using the projected future population and income levels, the model can then be used to predict future interzonal demand between A and B. The specific parameter values used in this example are given in the following table.
Table 5-3
Demand Parameter Values

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\alpha_0$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>$\alpha_3$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>$0.035 \times 10^{-6}$</td>
<td>0.88</td>
<td>1.00</td>
<td>0.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Demand is thus assumed to be unit-elastic with respect to the price variable. We will explore the sensitivity of the decision model to variations in this variable in Section 5.3.6.

We further assume that the demand parameter set $\phi = [\alpha, \beta]$ is the state variable which is subject to a degree of uncertainty. For the runs reported in this chapter, we have assumed $\alpha_0$, the variable which controls the magnitude of demand, is the uncertain state variable. Figure 5-4 shows schematically how demand will vary over time with an uncertain $\alpha_0$ and a fixed investment sequence.

5.2.4 The Socio-Economic Activity Shift Allocation Model

As we pointed out in Chapter 2.0, an investment in a facility at a particular point in time can affect the accessibility of the surrounding region. This in turn can result in shifts in the location patterns of the population and economic activity over time. As a
Figure 5-4
The Range of Demand over Time for Uncertain State Variable $\alpha_0$
consequence, flow volumes and ultimately, net benefits are also affected by these longer-run supply-demand dependencies. Demands for future periods therefore, are not only dependent on the supply available during this period but are a function of the sequence of investment in supply in previous periods.

In this study, we have modelled changes in demand over time due to changing accessibility by a rather simple activity shift model.\(^1\) The change in demand occurs in terms of a population shift from one region to another due to an increase or decrease in general accessibility.

The assumptions underlying this model are briefly summarized as follows:

1. The area-wide change in population can be specified exogenously as a percentage of the base year population.

2. The change in population in each zone for period \(t+1\) is a function of

   (a) the current population at time \(t\)

   (b) the percentage change in zonal accessibility, \(\Delta a_1\)

---

\(^1\)For obvious reasons of time and expense, we have not developed a full activity shift model that a more formal planning study might employ. Instead, we have used a simplified model of the activity system in terms of population and accessibility. This model is similar to one developed for a larger study effort. See Ruiter [5] for a more detailed discussion. In effect, we are incorporating this simpler model as a surrogate for a more realistic formulation. For descriptions of activity allocation models commonly used in transport planning, see for example Hansen [6], Irwin [7], Schlager [8] and Brand, et al [9].
The accessibility of a zone is defined in terms of a measure of attractiveness of each possible destination for trips from a zone, divided by a measure of the separation between the origin and destination zones. In more formal terms, the accessibility of zone \( i \) is defined as:

\[
A_i = \frac{K_a}{\alpha_0 i \alpha_1 Y_i \alpha_2 \sum_j V_{ij}}
\]

where \( K_a \) = a constant

\( P_i \) = the population of zone \( i \)

\( Y_i \) = the income of zone \( i \)

\( V_{ij} \) = the total number of trips from \( i \) to \( j \)

\( \alpha_0, \alpha_1, \alpha_2 \) = demand parameters

The linear form of this model is shown in Figure 5-5. The parameter, \( f \), determines the effect of accessibility on population changes. When \( f \) is small, the accessibility effect is minimal. For large \( f \), accessibility plays a major role in determining the location of population in the next period.

The general strategy for predicting population changes due to changed accessibility is outlined in Figure 5-6. Figure 5-6(a) shows how, given the activity system represented by the population and incomes for both time periods 1965 and 1970, an assignment of flows occurs and the measures of accessibility are computed. These measures are then used, along with exogenously predicted control totals on population, to predict the percentage zonal allocation of
Figure 5-5
Predicting Shifts in Population Due to Changing Accessibility
Figure 5-6
The General Strategy for Activity System Predictions
225
the regional growth of population, as a function of the network-related accessibility measure. The procedure is repeated in Figure 5-6(b), using the newly acquired 1975 data, along with the 1970 data to predict the 1985 socio-economic situation.

In Section 5.3.5.2, we use this simple model for a series of experiments to explore the effects on investment when demand is not only a function of the current investment, but a function of the sequence of investments to this point.

5.2.5 Equilibrium at a Stage and Net Benefits

There are a number of dimensions to the set of total benefits that occur when investment in a transport facility takes place. They include such externalities as increases in land values, changes in such measures as accessibility of a region, and changes in income and population due to an increase or decrease in this accessibility.

In the example presented here, although the program structure potentially has the capability for exploring these kinds of changes, we will concentrate solely on the primary cost and benefits attributable to the facility. Gross benefits are therefore defined as measured by the traditional economic measure of willingness-to-pay that we described in Chapter 2.0. This is equivalent to the price (including user tax payments) times the volume for each and every price up to the equilibrium value. Gross benefits for any facility $k$ are

\[ \text{Gross Benefits}_{k} = \sum_{p} (\text{Volume}_{k}(p) \times \text{Price}_{k}(p)) \]

---

1 This assumes the demand function can be interpreted as a marginal benefit function. See Wohl and Martin [10] for a derivation of the marginal benefit function.

226
therefore defined as the summation of the area under the demand curve, as shown in Figure 5-7.¹

The costs are comprised of fixed and variable costs. Fixed costs include facility investment and right-of-way costs. The variable costs include both operating costs and maintenance costs of the fixed facility. The operating cost is given by the perceived price at equilibrium, \( p_k \), minus the user taxes. The cost side is then given by

\[
C_k = (I_k + R_k) + (M_k + (p_k - X_k - X_k) v_k)
\]

where

- \( C_k \) = total costs
- \( I_k \) = facility investment cost
- \( R_k \) = right-of-way cost
- \( M_k \) = maintenance cost for link \( k \)
- \( X_k \) = average user tax
- \( v_k \) = flow volume on link \( k \)

¹This is equivalent to including consumer surplus (including the upper area of the curve) with the area \( p_k v_k \) in Figure 5-7.
Figure 5-7
The Measure of Gross Benefits: Willingness-to-Pay
The net benefits are computed as gross benefits minus costs or

\[ B_k = G_k - C_k \]

\[ = G_k - [I_k + R_k + M_k + (p_k - X_k) v_k] \]

Each of the costs and other characteristics of the 14 alternatives are shown in Table 5-4.

---

1. We have so far ignored the units in these equations. In order to put these costs on a comparable basis, we use the equivalent annual series approach of Section 4.7 for fixed costs. Since a facility may be constructed in any year and has continuing variable costs over its life, both benefits and costs must be discounted back to some base period using the discount factor \( d_t \) where:

\[ d_t = (1 + \rho)^{-t} \]

where \( t = \) year of benefit or cost occurrence

\( \rho = \) interest rate

229
<table>
<thead>
<tr>
<th>Alternative (k)</th>
<th>Capital Cost $I_k^{106}$</th>
<th>R.O.W. Cost $R_k^{106}$</th>
<th>Annual Maintenance Cost $M_k^{106/yr}$</th>
<th>User Taxes $X_k^{\text{veh-mile}}$</th>
<th>Life of Capital $L_m$ (yrs)</th>
<th>Life of R.O.W. $L_r$ (yrs)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.447</td>
<td>.0085</td>
<td>20</td>
<td>40</td>
<td>Existing at grade 4 lanes</td>
</tr>
<tr>
<td>2</td>
<td>1.000</td>
<td>0.354</td>
<td>0.0316</td>
<td>.0096</td>
<td>20</td>
<td>40</td>
<td>Improved existing, level 1</td>
</tr>
<tr>
<td>3</td>
<td>1.300</td>
<td>0.354</td>
<td>0.034</td>
<td>0.096</td>
<td>20</td>
<td>40</td>
<td>Improved existing, level 2</td>
</tr>
<tr>
<td>4</td>
<td>2.000</td>
<td>0.354</td>
<td>0.040</td>
<td>0.096</td>
<td>20</td>
<td>40</td>
<td>Improved existing, level 3</td>
</tr>
<tr>
<td>5</td>
<td>2.752</td>
<td>0.788</td>
<td>0.036</td>
<td>0.013</td>
<td>20</td>
<td>40</td>
<td>New 2 lane facility</td>
</tr>
<tr>
<td>6</td>
<td>2.752</td>
<td>1.100</td>
<td>0.036</td>
<td>0.013</td>
<td>20</td>
<td>40</td>
<td>New 2 lane plus r.o.w. for 4 lanes</td>
</tr>
<tr>
<td>7</td>
<td>5.690</td>
<td>1.300</td>
<td>0.050</td>
<td>0.013</td>
<td>20</td>
<td>40</td>
<td>New 4 lanes</td>
</tr>
<tr>
<td>8</td>
<td>5.600</td>
<td>1.100</td>
<td>0.050</td>
<td>0.013</td>
<td>20</td>
<td>40</td>
<td>Add 2 lanes to Alternative 3</td>
</tr>
<tr>
<td>9</td>
<td>5.100</td>
<td>1.100</td>
<td>0.050</td>
<td>0.013</td>
<td>20</td>
<td>40</td>
<td>Add 2 lanes to Alternative 4</td>
</tr>
<tr>
<td>10</td>
<td>3.050</td>
<td>0.900</td>
<td>0.040</td>
<td>0.013</td>
<td>20</td>
<td>40</td>
<td>New 2 lane facility</td>
</tr>
<tr>
<td>11</td>
<td>3.050</td>
<td>1.100</td>
<td>0.040</td>
<td>0.013</td>
<td>20</td>
<td>40</td>
<td>New 2 lane plus r.o.w. for 4 lanes</td>
</tr>
<tr>
<td>12</td>
<td>5.300</td>
<td>1.300</td>
<td>0.080</td>
<td>0.013</td>
<td>20</td>
<td>40</td>
<td>New 4 lanes</td>
</tr>
<tr>
<td>13</td>
<td>5.300</td>
<td>1.100</td>
<td>0.080</td>
<td>0.013</td>
<td>20</td>
<td>40</td>
<td>Add 2 lanes to Alternative 8</td>
</tr>
<tr>
<td>14</td>
<td>4.900</td>
<td>1.100</td>
<td>0.080</td>
<td>0.013</td>
<td>20</td>
<td>40</td>
<td>Add 2 lanes to Alternative 9</td>
</tr>
</tbody>
</table>

Note: Capital and right-of-way cost for 4 lane facilities are not the incremental costs of a 2 to 4 lane expansion, but the total cost.

Table 5-4 Technology Characteristics
230
5.3 Optimal Capacity Expansion under Conditions of Assumed Certainty and Uncertainty

The results and conclusions of these experiments can be divided roughly into three major categories, two of which we explore in detail in this chapter. The third is described in Chapter 6.0 after extending the basic sequential model.

The first category is the effect on investment levels of a deterministic versus a stochastic approach— for example, we want to determine whether or not staging or partial construction can offset the scale economies of immediate implementation, and how is this affected by uncertainty.

The second category concerns the efficiency of the search procedure— given the problem of staging investments over time with a probabilistic demand model that produces a large-scale extensive form decision tree, how well do the approximating procedures that we proposed in Chapter 4.0 reduce the problem size.

The results of these first two issues (investment policies under uncertainty and the results of incorporating pruning rules and terminal evaluation functions for the project design problem) are presented here.

The results that follow were produced using a series of computer programs that model both the transportation equilibrium flow procedure (MODEL) and the decision tree calculations (DECISN)
described in previous chapters. The examples chosen to demonstrate various points are variations of the example described in Section 5.2. These variations were achieved by changing fixed or variable costs, demand parameters, and so on in order to demonstrate a particular point. The programs themselves are extremely flexible and can be operated in either a deterministic or stochastic mode with as many or as few stages as desired and for a number of different discount rates. Most of the results that follow were with population growth rates of 2% or 3% per year and discount rates of 5% and 10%.

A comment is in order here on the limited number of stages and alternatives employed in most of the following examples. The number of stages were limited for the most part to either 2 or 3 because of the extreme amount of computation time involved. For example, with 3 state variables, 12 alternative plans to simulate at each stage, and 2 stages, a full tree search takes 21 seconds using an IBM 360/67 (with minimal output). For a 3 stage problem, the time increases to 3.5 minutes. At a cost of approximately $.10/second (it is actually something less than this at $325.00/hour), the cost for a 2 stage search is $2.10 and $21.00 for a 3 stage search. Needless to say, we have not yet attempted a full 4 stage stochastic tree search except for a more limited number of alternatives. For a similar reason, in order to demonstrate a particular point, we have sometimes limited the alternative space to either the existing alterna-
tive and its improvements (4 projects and 4 plans) or to the existing alternative and location G (12 projects and 24 distinct plans). (We will discuss the effects of variations in stages, alternatives and state variables on computation time further in Section 5.4.8.)

5.3.1 Comparison of Absolute versus Comparative Advantage

Before launching into the full 44-plan, stochastic tree search over a number of future investment periods, we conducted some simple constrained experiments to test:

(1) a number of time-staging issues described in previous chapters in a relatively uncluttered setting; and

(2) the flexibility of the program structure in its ability to constrain alternatives, stages, etc. with little variation in the basic data of the problem. This flexibility in constraining the problem is essentially equivalent to selective sampling of the tree at various decision points. This ability to reduce the solution space in a simple manner was one of the major objectives of the research and its value will become even more apparent when we turn to the full tree search of the larger problem.

This first set of experiments is designed to compare the effects of timing on investment choice for a deterministic demand structure.

Although it has been generally recognized by much of the literature on investment planning in other fields, transport investment planning has somehow failed to fully account for the potential benefits from multi-time period solutions to investment problems. As we stated in earlier sections, an early typical approach to investment planning was to select a design year for
some period in the future which represented some average future condition. The next step in the procedure was then to analyze several alternative investments, and then choose the investment which maximized whatever were the stated objectives initially set down for the problem. This produced a facility which was "best" for the design year but ignored initial period consequences. Most of the recent examples cited as project analysis or project design examples in transportation have at least extended this simple approach to include multiple time periods in the analysis [10,11]. These extensions then, now account for both initial and future consequences in evaluating different projects.

Unfortunately, this still misses the important point of a comparative versus absolute advantage, labelled by Marglin as the pure timing problem [12]. The rejection of a project for immediate construction with costs and benefits measured over a number of future time periods means only that the selected project has an absolute advantage over the rejected project. However, it ignores the fact that postponement of the rejected project, because of changing demands and/or the discount factor, may increase the benefits of that project such that it provides more total benefits than the originally accepted alternative. Thus the accepted project may have an absolute advantage over the rejected project; but by allowing investment postponement, it is possible that the rejected project will enjoy a comparative advantage over the initially
accepted project. The point is a simple one: the decision variable is now not only what project to invest in, but also when should investment take place.

Table 5-5 shows the results for the multiperiod model for the two possible cases of comparative and absolute advantage for a 6 alternative example. We constrain the alternatives to the existing alternative and its improvements ($\Lambda_1$, $\Lambda_2$, $\Lambda_3$, and $\Lambda_4$) and to $\Lambda_5$ and $\Lambda_9$ for location G (a new 2 lane facility or a new 4 lane).

For determining the project which enjoys absolute advantage, each investment alternative is constrained to a 0,1 case of acceptance or rejection for the total period. For comparative advantage, each alternative may also be delayed or postponed. The calculations are accomplished for the first case by iteratively evaluating each sequence, using what we will hereafter refer to as the forced sequence approach, and for the second case by simply constraining the state transformation matrix to allow only a single potential investment expansion for all alternatives. When evaluated in a full decision tree run, this second approach automatically compares alternatives for their comparative advantage of delayed construction. Demand is taken as deterministic with parameters $\alpha_k$ and $\beta_k$ equal to $0.035 \times 10^{-6}$ and 1.0, respectively.

The results indicate no difference between the plan sequence for the two cases for a 5% discount rate; plan $\Lambda_9$ is selected in both
Table 5-5
Comparison of Absolute vs Comparative Advantage

<table>
<thead>
<tr>
<th>Discount Rate</th>
<th>5%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 1</td>
<td>Stage 1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Sequences Showing Absolute Advantage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A_1) (A_1) (A_1)</td>
<td>-1.110</td>
<td>(A_1) (A_1) (A_1)</td>
</tr>
<tr>
<td>(A_2) (A_2) (A_2)</td>
<td>-0.0323</td>
<td>(A_2) (A_2) (A_2)</td>
</tr>
<tr>
<td>(A_3) (A_3) (A_3)</td>
<td>1.037</td>
<td>(A_3) (A_3) (A_3)</td>
</tr>
<tr>
<td>(A_4) (A_4) (A_4)</td>
<td>3.587</td>
<td>(A_4) (A_4) (A_4)</td>
</tr>
<tr>
<td>(A_5) (A_5) (A_5)</td>
<td>4.532</td>
<td>(A_5) (A_5) (A_5)</td>
</tr>
<tr>
<td>(A_9) (A_9) (A_9)</td>
<td>4.921</td>
<td>(A_9) (A_9) (A_9)</td>
</tr>
<tr>
<td>Optimal Sequence having Comparative Advantage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A_9) (A_9) (A_9)</td>
<td>4.921</td>
<td>(A_1) (A_9) (A_9)</td>
</tr>
</tbody>
</table>

Notes: Horizon 30 years
Stages 5-10-15 (Years)
ENPV - net present value

236
instances. At a 10% interest rate, however, the results indicate that for the absolute advantage case, alternative plan sequence \( \Lambda_5 - \Lambda_5 - \Lambda_5 \) is the best investment with net benefits of \$ -3.543 \times 10^5 \), or an increment of 61\% over the second best investment. By allowing the timing of investment to also vary, we observe a change in both the investment and the period from a \( \Lambda_5 - \Lambda_5 - \Lambda_5 \) sequence to a \( \Lambda_1 - \Lambda_9 - \Lambda_9 \) sequence, and a total improvement of net benefits of \$ .065 \times 10^6 \).

The point in time where investment plan \( \Lambda_9 \) becomes desirable is the point where the savings in interest costs on capital from delaying investment just balances the loss in benefits due to the delay. In this case, by allowing investment postponement, we observe a shift from a low capacity investment to a delayed high capacity alternative. The important point to note about these savings in unspent capital is that they can be used for other projects, or invested in some other form, to give an even higher return than the interest rate would indicate. If the discount rate measures the true opportunity cost of capital, then the savings indicated are correct.

5.3.2 Tradeoffs between Staging and Scale Economies for Deterministic Demands

One of the implicit benefits from adopting a multi-time period framework, in addition to the comparative advantage of delaying projects discussed in the previous section, is the fact that additional benefits can be achieved by also staging or partially constructing the facilities. The "indivisibility" or "lumpiness" of transport invest-
ments can be partially overcome using such a technique. The benefits from such a procedure are, again, savings in unused capital outlays for excess second stage capacity, as in the previous section. The losses associated with staging arise from the fact that the total costs for facilities built under staged construction will almost always be greater than if they had been built at a single point in time. In other words, there can be and often are, scale economies associated with building the larger capacity facilities all at once. Considering the staging periods properly, however, and accounting for the benefits accruing from alternative uses of the second stage capital in the interim, can result in total benefits which are greater than the benefits due to scale economy effects.

The alternatives available for investment in this example are now constrained to the existing alternative and location G. However, in place of the mutually exclusive alternative set of only \( \Lambda_5 \) and \( \Lambda_9 \) (the two and four lane facilities), \( \Lambda_5 \) can also be expanded to a four lane facility (alternative \( \Lambda_7 \)) at an overall total cost slightly more than building the full four lanes immediately.

The results of this run for a demand level of \( \alpha_2 = 0.035 \times 10^{-5} \) show a shift in the investment pattern to \( \Lambda_5-\Lambda_5-\Lambda_7 \) which is more beneficial than simply building the full capacity or alternative \( \Lambda_9-\Lambda_9-\Lambda_9 \) all at once or even the postponed high capacity sequence, \( \Lambda_1-\Lambda_9-\Lambda_9 \).
Of course, there is nothing unusual in these results, or the results of the previous section, nor is the approach especially unique. Tradeoffs between staging policies and scale economies have been explored by a great many authors, as we indicated in Chapter 2.0. Some of the major differences between these approaches and ours, however, is:

(1) the specification of aggregate capacity

(2) the use of supply-sensitive demand functions, and

(3) the need for a two phase process of flow equilibrium calculation and evaluation.

A fourth major difference and perhaps the most important one is the extension to a sequential decision framework. Traditional comparisons of staging techniques and economies of scale have considered tradeoffs between benefits from immediate investment (the cost decrease due to scale of a 4 lane immediate investment, over a staged, 4 lane facility) and the gains from using the capital savings from a staged sequence plan (the savings of a delayed second stage investment) for a fixed and deterministic set of demands. The inclusion of first, uncertainty in the demand expected to use the facility (the case of static uncertainty), and second, in the nature of the decision process (making sequential conditional decisions) should have quite different effects on these tradeoffs.
A priori, we would expect a stage construction policy to be less beneficial than immediate investment for the non-sequential static uncertainty case, all else held constant, for the normal case of higher losses from under-investment discussed in the last section. That is, the gains from staging will be more than offset by losses expected if demands are much higher than expected. On the other hand, staging techniques should prove extremely beneficial once a sequential dynamic uncertainty framework is included.

If demands are uncertain, the loss from a "wrong" decision can be lessened by only partially investing in a particular facility. The benefits from waiting outweigh those attributable to scale economies. If demand does not materialize, there will be considerable savings due to a revised second period decision. If demand is higher than expected in the first period, investment can still proceed in the larger facility expansion with only minor losses from a one period wait and of course, from not taking advantage of scale economies.

In summary, the decision to stage, or not to stage construction, will be a function of the particular problem in question: the degree of scale economies present, the growth in demand, the discount rate, and so on. The question of staging and scale economy tradeoffs is one thing in a deterministic world, however, and quite another in the real environment of sequential decision making when demands are uncertain and decisions are continually being revised [13]. But
before we explore timing and staging issues, and the full sequential problem under uncertainty, we first turn to explore a simpler problem involving stochastic demands and one additional feature of staging policies, i.e., building adaptability into the technology itself.

5.3.3 **Flexibility, Adaptability and the Staging Problem under Uncertainty**

An additional factor to consider when staging investments, especially if demands are to some degree uncertain, is to provide for flexibility in the investment itself, as we outlined briefly in previous chapters.

There are any number of ways of doing this: in the typical inventory control problem, flexibility is provided for by building up inventories to cope with unexpected chance situations of extreme demand levels. In this case, flexibility is the excess capacity that is available to offset unforeseen demands. The reason for carrying more supply than the average demand requires, is that the loss associated with an exhausted supply is considerably more than the loss of storing the excess inventory.

A second kind of flexibility is to use a technology which is not optimal for any specific output but relatively good for a wide range of outputs, as we outlined in Chapter 1.0. The benefits of such an approach are obvious if output varies either randomly over a wide range, or operates at different output levels because of peaking or some other factor. A good example of this kind of flexibility is the 727 QC airplane designed to operate as a passenger carrying aircraft during the day and to quickly convert to a cargo
aircraft for night operation [14]. In this case, the aircraft is probably not optimal for either operation but it is relatively efficient for each kind of output and overall, more desirable than either of the technologies which are optimal for their respective output levels, but undesirable for other output levels.

Another example of technology flexibility is given by Heflebower [15] for pipeline capacity expansion. If throughput volume is an uncertain factor, constructing a larger diameter pipe with fewer and more widely spaced pumping stations results in a service which costs little more than a smaller capacity pipe operating at its optimal operating point. If throughput is greater than expected, the capacity of the line can be expanded with little expense - it merely involves adding more pumping stations. The factor which allows this "flexibility" at little cost, is the scale economies associated with pipeline technology.

The analogy in the highway case is the use of an oversized facility with widely spaced interchanges, or constructed to fairly low design standards initially. Increases in the expected volume can then be handled by improving the design characteristics, or adding additional interchanges.

A third kind of flexibility is better termed as adaptability. This refers to the ability to change relatively fixed resources to adapt to changing conditions. Massé [16] gives the examples of temporary pre-fabricated houses which can be moved easily if condi-
tions warrant a change, as opposed to a permanent structure. Other examples may involve simply providing the capability of remodelling the interior of a building, i.e., providing movable partitions instead of fixed walls.

There are many other examples of flexibility¹ - most of the literature refers to it as an intangible quantity, something to consider after all costs are evaluated. For example, Hanssmann [17] states:

"Cost is a tangible decision criterion. At the same time it is clear that cost cannot be the only consideration. Among other things, the military executives are interested in the flexibility of the fleet, in order to be prepared for emergencies that differ from the "normal" type of mission. Within the framework of this study flexibility was an intangible consideration."

We contend that flexibility (in the general sense) is not an intangible quantity but that it is inherent in the technology and can be directly evaluated within a stochastic framework. Moreover, flexibility and adaptability are not free goods - in order to buy flexibility, we must give up some other resource. On the other hand, the benefits of buying flexibility are a higher overall expected efficiency. Additionally, there is a tradeoff between flexibility and adaptability - the more you have of one, the less you need of another.

¹A good example of flexibility in new technology is the dual-mode system which is probably not as efficient as rail rapid transit for hauling high volumes of people in high density areas nor as efficient as the auto for the low density areas - but over all, it provides a service which is passably efficient in both regions.
In the single link case, we can introduce any number of ways of providing flexibility or adaptability. Providing traffic signals, interchanges, channelization, etc. is one way of improving the throughput of an existing facility. Buying right-of-way for expansion of a new facility from a 2 to 4 lane facility is another way of providing adaptability of investment.

To explore these general concepts, we introduce in this section the problem of the value of pre-investment for the single link case. We use Massé's model adapted to the highway case as a basis for this example [16].

For this example, we assume that the initial decision is between an "adaptable" and an "unadaptable" 2 lane facility. The adaptability factor is the provision of right-of-way for future expansion. Using the project design example of previous sections, we constrain the initial investments to the following:

(1) \( \Lambda_6 \): to build a 2 lane facility now, with right-of-way for a 4 lane facility and (2) \( \Lambda_5 \): a 2 lane with only enough right-of-way for 2 lanes. Both facilities offer the same services and differ only in investment and maintenance costs. These costs are represented by \( c(\Lambda_5) \) and \( c(\Lambda_6) \) where \( c(\Lambda_6) > c(\Lambda_5) \).\(^1\)

\( c_t(\Lambda_k) \) thus represents the cost at stage \( t \) of alternative \( \Lambda_k \).

---

\(^1\) We will couch our discussion in terms of costs alone; however, in our experiments, we actually use the consumer surplus measure of benefits as well as cost as described earlier.
We further restrict our example to two stages \((t = 1 \text{ or } 2)\).

Let us now assume that the decision has been made for stage 1 (i.e. it is either \(\Lambda_5\) or \(\Lambda_6\)) and we now want to determine for stage 2 whether to expand capacity to a 4 lane or not. The costs of operation of the facilities are the supply functions for the 2 and 4 lane facilities shown schematically in Figure 5-8 (a). The cost of adapting or not adapting depends on the demand level in the second stage, which in this example is a random variable. The loss associated with not adapting can be defined in part, as the difference, \(a(v)\), between the two supply functions for any volume, \(v_2\), shown in Figure 5-8 (a).

If we redraw this loss \(a(v)\) for all volume levels on a new diagram in Figure 5-8 (b), we can then add to this diagram the marginal cost of expansion to a 4 lane for the two alternatives \(\Lambda_5\) and \(\Lambda_6\). This is represented by \(c_2(\Lambda_7)\) and \(c_2(\Lambda_8)\), the additional costs of expanding to a 4 lane facility for \(\Lambda_5\) and \(\Lambda_6\) alternatives, respectively. Looking at these curves on the same graph, we can identify the break points, \(x\) and \(y\). We now have 3 zones of interest for the stage 2 volume \(v_2\): \(v_2 < x\), \(x < v_2 < y\), and \(v_2 < y\).

Case 1: \(v_2 < x\)

In this zone, it will be desirable not to expand to the 4 lane facility, no matter what the initial decision had been. The loss associated with operating a 2 lane facility is less than either \(c_2(\Lambda_7)\) or \(c_2(\Lambda_8)\).
Figure 5-8
The Adaptability of Technology
246
Case 2: $x < v_2 < y$

If the initial decision had been the adaptable alternative, $\Lambda_6$, it would pay to expand to the 4 lane facility ($a(v) < c_2(\Lambda_8)$).

If the original decision was the unadaptable alternative, it would pay to incur the loss $a(v)$ and not expand.

Case 3: $y < v_2$

For the third and final zone of interest, it would pay to expand or adapt to a 4 lane facility in either case, i.e. no matter what the initial decision since $a(v_2) < \text{either } c_2(\Lambda_7) \text{ or } c_2(\Lambda_8)$.

The value of adaptation should now be obvious. It is not a free good nor is it desirable in all cases. It can offset an initially higher fixed cost which produces two advantages. First, it lowers the cost of adaptation to a higher capacity facility in the zone, $y < v_2$, where adaptation is carried out in any case, and it broadens the zone where adaptation would be profitable, $x < v_2 < y$.

The above discussion has been concerned with the decision at stage 2 when the decision at stage 1 is known. We now turn to determining that initial decision. We use as an example a restricted version of the single link data of Section 5.2. We limit the alternatives through the state transformation matrix to (1) $\Lambda_1$, the existing alternative, (2) $\Lambda_5$ and $\Lambda_7$, the 2 to 4 lane unadaptable expansion sequence, and (3) $\Lambda_6$-$\Lambda_8$, the adaptable 2 to 4 lane expansion. The supply functions are as defined in Section 5.2. Population growth is 2%, the discount rate is at 10%, and the horizon is 30 years with
two stages of 15 years in length. Demand can take on three values for each stage with the state variable levels taking on values, \( \alpha_k \), of \( 0.035 \times 10^{-6} \), \( 0.050 \times 10^{-6} \) and \( 0.065 \times 10^{-6} \) with equal probability \( 0.333 \) for both stages. Investment and right-of-way costs are given in Table 5-6(a). Results of the decision tree calculations for these conditions are shown in Table 5-6(b). They indicate that the expected cost of adaptation is not worth the benefits it would produce for the \( \alpha_2 \) and \( \alpha_3 \) state variable spines, i.e. in stage 2, we are in zone \( \nu_2 < \gamma \), and it pays to adapt to the larger facility, \( \Lambda_7 \), but over all, based on the demands for both periods, it will not pay to incur the added pre-investment cost of the \( \Lambda_6 - \Lambda_7 \) sequence.

In order to see when the adaptable alternative becomes desirable, we then varied the "cost of adaptation" for the \( \Lambda_6 - \Lambda_8 \) sequence between the two extremes \( 0.788 \) to \( 1.3 \) of the \( \Lambda_5 - \Lambda_7 \) sequence. Results are shown in Table 5-7. For a r.o.w. cost of \$ \( 0.900 \times 10^6 \) for \( \Lambda_6 \) for all 4 lanes, the sequence \( \Lambda_5 - \Lambda_7 \) is desirable for state variable spines \( \alpha_2 \) and \( \alpha_3 \). For a r.o.w. cost of \$ \( 0.870 \times 10^6 \), the alternative sequence changes to \( \Lambda_6 - \Lambda_8 \) for the \( \alpha_2 \) and \( \alpha_3 \) spines. (Note that for \( \alpha_1 \), the investment sequence is \( \Lambda_1 - \Lambda_5 \) in both cases.)

Thus, for a cost of building adaptability into the 2 lane investment of less than \$ \( 0.082 \times 10^6 \), \( (0.870 - 0.788) \), the adaptable alternative is more desirable for 2 of the three state variable levels, and is also chosen as the best investment.
Table 5-6 (a)
Investment and R.O.W. Costs for the Adaptable Technology Example

<table>
<thead>
<tr>
<th>Alternative</th>
<th>Incremental Fixed Cost of Expansion (x10^6)</th>
<th>R.O.W. Cost (x10^6)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Unadaptable Λ₅</td>
<td>2.752</td>
<td>.788</td>
<td>2 lane</td>
</tr>
<tr>
<td>Λ₇</td>
<td>2.848</td>
<td>.522</td>
<td>4 lane</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Total Cost</td>
</tr>
<tr>
<td>2 Adaptable Λ₆</td>
<td>2.752</td>
<td>1.1</td>
<td>2 lane with r.o.w. for 4</td>
</tr>
<tr>
<td>Λ₈</td>
<td>2.848</td>
<td>0.</td>
<td>4 lane</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Total Cost</td>
</tr>
</tbody>
</table>

Table 5-6 (b)
Results

<table>
<thead>
<tr>
<th>Demand</th>
<th>Investment Sequence</th>
<th>Net Present Value (x10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 1</td>
<td>Stage 2</td>
</tr>
<tr>
<td>α₁</td>
<td>Λ₁</td>
<td>Λ₅</td>
</tr>
<tr>
<td>α₂</td>
<td>Λ₅</td>
<td>Λ₇</td>
</tr>
<tr>
<td>α₃</td>
<td>Λ₅</td>
<td>Λ₇</td>
</tr>
<tr>
<td>Stochastic</td>
<td>Λ₅</td>
<td>Λ₇</td>
</tr>
</tbody>
</table>
Table 5-7
Changes in Investment Sequences for Variations in the Cost of Adaptation

<table>
<thead>
<tr>
<th>R.O.W. Cost Variation for Λ₆ (x10⁶)</th>
<th>Demand Level</th>
<th>Investment Sequence</th>
<th>Net Present Value (x10⁶)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α₁</td>
<td>Λ₁    Λ₅</td>
<td>- .647</td>
</tr>
<tr>
<td>.900</td>
<td>α₂</td>
<td>Λ₅    Λ₇</td>
<td>.482</td>
</tr>
<tr>
<td></td>
<td>α₃</td>
<td>Λ₅    Λ₇</td>
<td>2.22</td>
</tr>
<tr>
<td></td>
<td>Stochastic</td>
<td>Λ₅    Λ₇</td>
<td>.485</td>
</tr>
<tr>
<td>.870</td>
<td>α₁</td>
<td>Λ₁    Λ₅</td>
<td>- .647</td>
</tr>
<tr>
<td></td>
<td>α₂</td>
<td>Λ₆    Λ₈</td>
<td>.494</td>
</tr>
<tr>
<td></td>
<td>α₃</td>
<td>Λ₆    Λ₈</td>
<td>2.23</td>
</tr>
<tr>
<td></td>
<td>Stochastic</td>
<td>Λ₆    Λ₈</td>
<td>.479</td>
</tr>
</tbody>
</table>
sequence as an unconditional sequence under uncertainty. For an
alternative within this cost range, the expected future benefits
of providing an adaptable alternative outweigh the initial cost of
providing that adaptation.

In the present model, conditional adaptation is contained
in the alternative selected. At the second stage, the choice is
predetermined by the initial choice; we are no longer free to adapt
or not adapt. The decision as to whether expansion occurs or not is
determined by the expected value of the state variable at that stage.

The question now arises as to the optimum amount of adaptability
to build into a facility. This will depend primarily on the
cost of providing adaptation and the expected benefits attributable
to it. Additionally, the example could also now easily be expanded
to include more than one kind of flexibility and adaptability -
providing interchanges which affect the supply functions, building
alternatives to different design standards and so on.

Clearly, the concepts of flexibility and adaptability need
much more investigation. We present this introductory discussion here
to demonstrate its direct relationship to the stochastic investment
problem. Their advantage in the decision tree context is that the
sequential decision model, given the probabilities and cost func-
tions, can evaluate the optimum amount of flexibility and adaptability
required automatically. We will return to these concepts in the
following sections and in Chapter 6.0 in terms of the decision tree
strategy calculations.

251
5.3.4 Comparison of Investment Sequences for Deterministic and Uncertain Demand Parameters

Turning from the constrained examples of the previous three sections, we next consider deterministic and stochastic versions of the model using the full set of staged and unstaged, adaptable and unadaptable alternatives for the existing location, 0, and location G. There are therefore 24 alternative plans to consider.¹

A logical first step in comparing the results of a deterministic model to a stochastic model is to look at variations in the investment sequence for different demand levels. Selecting high and low values of the demand parameter $\alpha_k$ and running a full global search over each level separately for all stages results in the sequences shown in Table 5-8, for 3% and 2% population growth rates. For the 3% growth rate, and 5% interest rate, the low demand spine optimal sequence, $S_1^*$, contains alternative $\Lambda_1$ (the existing alternative) in period 1 (year 5) and alternative $\Lambda_2$ (the 1st level improvement to the existing alternative) in year 15. The mean demand spine is as in a previous section (sequence $\Lambda_5^*-\Lambda_5^*-\Lambda_5^*$). For high demands, $\alpha_3$, the optimal sequence $S_3^*$ is to invest immediately and continue with alternative $\Lambda_9$ (the 4 lane facility). This now gives us the optimal investment sequence for each demand level. The optimal alternatives and demand levels for

---

¹To be consistent with following examples, we have also now limited the period of interest to 15 years in length. In this and following sections, we will be dealing with 2 and 3 stage models with stage lengths of 5 and 10, or 5, 5, and 5 years.
Table 5-8

Optimal Investment Sequence for Selected Demand Level Spines

<table>
<thead>
<tr>
<th>Growth Rate</th>
<th>Demand Level $(x10^{-6})$</th>
<th>Investment Sequence</th>
<th>Net Present Value $(x10^6)$</th>
<th>Investment Sequence</th>
<th>Net Present Value $(x10^6)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
<td>10%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3%</td>
<td>$\alpha_1 = .015$</td>
<td>$\lambda_1 \lambda_1 \lambda_2$</td>
<td>-1.604</td>
<td>$\lambda_1 \lambda_1 \lambda_1$</td>
<td>-1.179</td>
</tr>
<tr>
<td></td>
<td>$\alpha_2 = .035$</td>
<td>$\lambda_5 \lambda_5 \lambda_5$</td>
<td>2.782</td>
<td>$\lambda_1 \lambda_2 \lambda_3$</td>
<td>- .446</td>
</tr>
<tr>
<td></td>
<td>$\alpha_3 = .050$</td>
<td>$\lambda_9 \lambda_9 \lambda_9$</td>
<td>10.07</td>
<td>$\lambda_3 \lambda_4 \lambda_4$</td>
<td>1.054</td>
</tr>
<tr>
<td>2%</td>
<td>$\alpha_1 = .015$</td>
<td>$\lambda_1 \lambda_1 \lambda_1$</td>
<td>-1.876</td>
<td>$\lambda_1 \lambda_1 \lambda_1$</td>
<td>-1.109</td>
</tr>
<tr>
<td></td>
<td>$\alpha_2 = .035$</td>
<td>$\lambda_2 \lambda_2 \lambda_3$</td>
<td>.764</td>
<td>$\lambda_1 \lambda_2 \lambda_2$</td>
<td>- .862</td>
</tr>
<tr>
<td></td>
<td>$\alpha_3 = .050$</td>
<td>$\lambda_5 \lambda_3 \lambda_5$</td>
<td>4.732</td>
<td>$\lambda_2 \lambda_3 \lambda_4$</td>
<td>- .012</td>
</tr>
</tbody>
</table>
$\alpha_1$ and $\alpha_3$ are also shown schematically in Figure 5-9 (a) in the traditional "demand and capacity vs. time" graph of the capacity expansion literature. Capacity in the transport case is a rather ill-defined term. It has any number of definitions ranging from free-speed volume (practical capacity) to maximum possible throughput (possible capacity). The exact definition is unimportant for our discussion except by the way it affects the cost function.

For purposes of the schematic, we will represent each alternative by its minimum average cost point, although this is considerably less than the volume that can be throughput.

Before running the full stochastic tree search, it would be useful to run each of the optimal alternative sequences for each demand level against each of the other demand levels to see how well the alternatives fare for $\alpha$ values other than the one they are optimal for. This entails six additional runs using the forced sequence mode; $S_1^*$ against $\alpha_2, \alpha_3$, $S_2^*$ against $\alpha_1, \alpha_3$ and $S_3^*$ against $\alpha_1, \alpha_2$. The results are contained in Table 5-9 and shown schematically in Figure 5-9 (b) for $\alpha_3$ only. Note from the figure that although we are evaluating each sequence against other $\alpha$'s, because the uncertain state variable is the demand parameter $\alpha$, and because demand is a function of the alternative selected, that there are different volumes occurring for each sequence when evaluated using a specific state variable spine, $\alpha_k$. Thus, not only do we generate an optimal sequence of investments when we search over a specific state parameter, but we produce an optimal volume sequence, or more
Optimal Sequences and Volumes for High and Low Demand Spines

Optimal and Non-Optimal Volumes for the $\alpha_3$ Spine

Figure 5-9
Optimal Investment Sequences for Alternative Demand Spines
255
Table 5-9
Results of Optimal Sequences at other Demand Levels

<table>
<thead>
<tr>
<th>Stage</th>
<th>Optimal Sequence</th>
<th>State Variable Level (x10^{-6})</th>
<th>Volume NPV (x10^6)</th>
<th>Volume NPV (x10^6)</th>
<th>Volume NPV (x10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\alpha_1 = 0.015$</td>
<td>$\alpha_2 = 0.030$</td>
<td>$\alpha_3 = 0.050$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$\Lambda_1$</td>
<td>230</td>
<td>490</td>
<td>659</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$S_1^*$ $\Lambda_1$</td>
<td>292</td>
<td>609</td>
<td>1363</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$\Lambda_2$</td>
<td>468</td>
<td>934</td>
<td>1633</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\alpha_1 = 0.015$</td>
<td>$\alpha_2 = 0.030$</td>
<td>$\alpha_3 = 0.050$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$\Lambda_5$</td>
<td>433</td>
<td>863</td>
<td>1127</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$S_2^*$ $\Lambda_5$</td>
<td>540</td>
<td>1050</td>
<td>1363</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$\Lambda_5$</td>
<td>667</td>
<td>1274</td>
<td>1633</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$\Lambda_9$</td>
<td>479</td>
<td>1012</td>
<td>1365</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$S_3^*$ $\Lambda_9$</td>
<td>605</td>
<td>1262</td>
<td>1680</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$\Lambda_9$</td>
<td>764</td>
<td>1556</td>
<td>2063</td>
<td></td>
</tr>
</tbody>
</table>

Notes:

NPV - Net Present Value

256
correctly, a unique set of volumes associated with the optimal investment sequence. The variation between each optimal sequence and the two non-optimal sequences for each $\alpha$ are obviously not very significant. This suggests that the investment to be selected from a full stochastic tree search will not be significantly different from one of the three sequences already chosen.

Let us now explore the effects of uncertainty on the initial investment decision. Recall that by the non-sequential static uncertainty problem definition of Chapter 2.0, we imply that the probability distribution is unchanging over time and we intend to make all decisions now — there will be no updating or changing of decisions in a later period. With that definition, we can then compare each of the alternative sequences for all possible demand levels using the full stochastic tree evaluation. This is accomplished by again using the forced sequence approach of previous sections in an iterative manner but now allowing demand to be a random variable. For 3% growth, 5% interest rates of Table 5–8, the initial investment for assumed certainty (i.e. $\alpha_2 = 0.035$) is alternative $\Lambda_5$ (adding 2 lanes) with an expected value of $2.782 \times 10^6$. For the non-sequential static uncertainty case, evaluating each alternative sequence using the full multi-period expected value approach results in a best sequence $\Lambda_2 - \Lambda_3 - \Lambda_4$ with an expected value of $2.183$. Thus, in the face of uncertainty, the sequence chosen is different from the one chosen by the deterministic model. In this case, the expected value is less than the value predicted using the mean state variable level.
In general, it can be either higher or lower depending on the losses associated with the high and low state variable levels.

5.3.5 The Effect of a Sequential vs a Non-Sequential Decision Procedure

With the previous discussion of high and low spines, non-sequential static uncertainty, and so on, as background, we can now turn to compare the effects of a full conditional tree search or the true sequential decision process described in Section 2.4, with the stochastic and deterministic results of the previous section.

Let us now assume that for the problem of Section 5.3.4, which suggests investing initially in alternative $A_5$ (the 2 lane expansion) for the mean demand level and in alternative $A_2$ for the non-sequential stochastic solution, we check the results of demand that have actually occurred after the stage 1 investment has taken place.

Given that the demand that actually occurs is different from the mean value, and that we may now revise our second period decision, we would expect, in general, that the optimal investment vector from stage 2 on would differ from the one initially chosen based on an expected demand, deterministic approach. Similarly, in the case of the non-sequential static uncertainty solution, if given the opportunity, we would also expect to revise the investment decision for the second and subsequent periods, instead of continuing with the original investment sequence.

258
Note, however, that we have come to a major crossroad in the philosophy of planning as presented in Section 2.3. The case of revising our decisions at the end of the first period and using a deterministic or non-sequential static uncertainty approach for predicting a new optimal investment sequence, observing results, again predicting a new sequence, and so on, will lead in general, to a different sequence of investments than the one that a true sequential model would select. This is because the former approach ignores at the time of prediction the fact that we know we will be revising our decisions at the end of the period. However, the net benefits of the actual revised investment pattern chosen by a sequential application of the non-sequential stochastic approach should certainly be higher than continuing with the original investment vector.

By including the fact that we do know revision will occur at each and every period, however, the procedure becomes a true sequential decision procedure and the initial investment pattern should again be different. In other words, if in the initial simulation, we consider the fact that we will be revising our decisions at the end of the first period, the investment decision will differ from that predicted either by the assumed certainty or the non-sequential stochastic case and the true expected net benefits from that decision will be higher still. That is not to say we ignore future consequences and decide myopically period by period what the investment should be. Rather, decisions now become conditional on previous period outcomes — our approach is now aimed at determining
strategies instead of sequences of investment.

With this procedure, assuming no additional uncertainty over our demand formulations, no expected change in goals and so on, theoretically we need only perform one simulation for all time. The decision analysis algorithm of Chapter 3.0 will perform the tree climbing or averaging-out and folding-back procedure and save the conditional investment decisions for all cases of demand occurrences. The initial investment is then the first investment of the best strategy chosen from among the complete set of strategies. This best strategy has implicitly included the benefits from knowing a priori that the decision will be revised at the second and subsequent periods.

In practice, the tree will be very large (we will not be able to compute all possible conditional decisions), our models are not known for certain,\(^1\) and our goals undergo revision and additions as the future is revealed to us. Therefore, even this procedure must be repeated using an approximate strategy at each stage. Even so, the decisions will be, in general, much improved over either of the cases of deterministic demands, non-sequential static uncertainty with no revision, or non-sequential static

---

\(^1\)Note the fact that we distinguish between model certainty, or uncertainty, and an underlying random nature to the demand structure itself. In practice, this merely means that we have some degree of uncertainty over the true underlying random nature of demand, i.e. our probability estimates also have a degree of uncertainty (see Chapter 6.0).
uncertainty with revision.

In order to test the model as a sequential decision process, we have performed a series of experiments with and without the activity-shift model of Section 5.2.4 operative. Demand now varies over its full range, by way of selected discrete levels of the state variable for all periods. The results of these two sets of experiments are contained in the following two sections.

5.3.5.1 Results for the Case of Long-Run Supply-Demand Independence

Table 5-10 compares the results of four runs using the tree-climbing sequential decision procedure of Chapter 3.0 and the former cases of assumed certainty and non-sequential static uncertainty. Note that the pattern of investment predicted with the sequential model is identical to the non-sequential stochastic results, and for Case 1 (2% growth), even with the pattern chosen using deterministic assumptions.

It is here that we can now fully appreciate the distinction between the traditional models of inventory control and the restricted project design example of this section (restricted in the sense of no long run supply-demand interdependence and no correlation between demand levels of contiguous stages). Briefly, the basic inventory control model has a loss function similar to the average cost function of our transport example. It is made up of (1) a loss associated with warehousing costs of storing inventory not used from one period to the next, i.e. a loss associated with overcapacity and (2) a
Table 5-10
Comparison of Deterministic, Non-Sequential and Sequential Solutions for the Two Stage Model

<table>
<thead>
<tr>
<th>Growth Rate</th>
<th>Discount Rate</th>
<th>Deterministic Solution</th>
<th>Non-Sequential</th>
<th>Sequential</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Investment Sequence</td>
<td>ENPV Stage</td>
<td>ENPV Stage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stage 1 2 (x10^6)</td>
<td>Stage 1 2</td>
<td>Stage 1 2</td>
</tr>
<tr>
<td>2%</td>
<td>5%</td>
<td>( \Lambda_2 \ \Lambda_3 \ 0.755 )</td>
<td>( \Lambda_2 \ \Lambda_3 \ 0.393 )</td>
<td>( \Lambda_2 \ \Lambda_3 \ 0.393 )</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>( \Lambda_1 \ \Lambda_2 \ 0.854 )</td>
<td>( \Lambda_1 \ \Lambda_2 \ -0.941 )</td>
<td>( \Lambda_1 \ \Lambda_2 \ -0.941 )</td>
</tr>
<tr>
<td>3%</td>
<td>5%</td>
<td>( \Lambda_5 \ \Lambda_5 \ 2.721 )</td>
<td>( \Lambda_2 \ \Lambda_4 \ 2.209 )</td>
<td>( \Lambda_2 \ \Lambda_4 \ 2.209 )</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>( \Lambda_1 \ \Lambda_3 \ -0.443 )</td>
<td>( \Lambda_1 \ \Lambda_3 \ -0.555 )</td>
<td>( \Lambda_1 \ \Lambda_3 \ -0.555 )</td>
</tr>
</tbody>
</table>

Note:

NPV - net present value

ENPV - expected net present value
penalty loss if demand greatly exceeds the available supply in any period known as a "stockout" cost; i.e. this is generally thought of as a loss of goodwill associated with undercapacity.¹ The loss associated with any supply decision at any stage depends on two items of information:

(1) the action or supply decision and the demand level at that stage, and

(2) the inventory level from the previous stage.

Thus, the expected loss for any decision, although independent from the sequence of previous supply-demand transactions, is dependent on the net inventory of the previous stage. Solutions to the general inventory problem take advantage of this separability and the recursive function approach of dynamic programming. The result is a true sequential decision process, which produces a series of conditional decisions dependent on the level of demand experienced in each period.

On the other hand, the transport time-staging problem as we have defined it to this point, does not depend on the sequence of investment prior to the current stage.² All we need to know is

¹For a revision of this concept of undercapacity loss, see Schwartz [18].

²This of course, assumes that the capital cost of investment is not a function of time. This is not too unlikely a situation. Maintenance as well as capital costs are likely to increase with the passage of time. We have programmed this capability into our basic model but have not yet extensively tested these effects on the investment decision.
the action at the current stage and the state variable level, \( \phi_k \). The reason for this independence is as follows: with no long-run supply demand dependence (i.e. demands are not affected by the path of capacity investment from previous stages) nor correlations between demands (i.e. if demand is high (or low) in the previous period, it is more likely to be high (or low) in the following period), the choice mechanism for the optimal alternative will choose the best alternative based on the expected value calculations for the \( n^{th} \) stage demands for each possible previous period investment.

This is essentially identical to the result of Cole's work [19]. His model is less constrained, however, since he assumes the stages where capacity expansion take place are separable in terms of fixed costs also. That is, there is no physical constraint from a previous period. Therefore his solution technique is to proceed in a stage by stage forward-seeking stepwise manner, i.e. he calculates the best investment under uncertainty for the first stage (independent of all future decisions) and then steps to the next stage. Basically, the procedure assumes stage decisions have no life associated with them. A decision to invest in capacity now has no constraint on either future alternatives or future costs.

Here we also note the difference between tracking adaptive and learning adaptive. With a stationary probability distribution, there is no tracking adaptive solution for the transport example as we have defined it. That is, the conditional decision tree reduces to a single sequence because, unlike the inventory problem, there is
no loss function conditional on the last stage. Therefore, the strategy (conditional decision) reduces to an unconditional sequence. At the second stage, we can ignore observing demand and revising our decision - assuming our model, etc. is an accurate representation of the real world - the decision remains the same no matter what the outcome. However, if there is long-run dependence or if there is correlation between demands, then a tracking adaptive solution becomes a reality as we shall see in the following section.

If we further assume that the underlying distribution over the demand parameter set is unknown and employ a learning mechanism, then we again have a conditional decision tree - but now because we are learning about an unknown distribution. We shall explore this extensively in the following chapter.

With this distinction between inventory capacity expansion models and the transport problem clearly spelled out, let us now turn to incorporating the long-run activity shift model and observe the effects on investment patterns.

5.3.5.2 Incorporating the Accessibility-Related Activity Shift Model

The primary feature of all previous runs to this point has been the absence of a longer-run supply dependent activity shift model. All solutions could have been reduced in terms of number of calculations and time by incorporating a backwards-recursive dynamic programming formulation similar to the one described in Section 3.3.1. Without supply-dependence over the longer run, the necessary condition of sep-
arability holds, and duplicated computations can be greatly reduced.

Such an independence between transportation and socio-economic activities in the long-run, however, is not a valid assumption, if we can use reported existing capacity expansions and shifts in population, commercial and industrial locations as examples. Clearly, the effect is not easy to predict. Considerable effort has been devoted to both urban and regional planning models in an attempt to capture these long-run dependencies (termed land-use modelling in the urban models) with very little reported in terms of validation. Nonetheless, for purposes of this study, we will assume that such dependencies exist and can be modelled in a fairly straightforward fashion, and focus our attention on the effects of such dependencies on investment policies under uncertainty.

In order to test these effects, we used the simple activity shift model described in Section 5.2.4 as a surrogate for a more complete socio-economic growth allocation model. By varying the $f$ factor we can change the amount that "accessibility" affects the patterns of population and economic factors of a multi-zone region.

Results of a series of experiments on the effects of accessibility are shown in Table 5-11 for variations in the accessibility factor $f$. Since our initial example is a simple two zone model, we might first expect that increased capacity between these zones should have a similar (and therefore cancelling) effect on population shift - the result being no shift in allocation due to an increase in capacity. The reason that this is invalid is found in our basic definition of
Table 5-11
The Effect of Accessibility on the Investment Sequence

<table>
<thead>
<tr>
<th>Accessibility Factor, f</th>
<th>Demand Level</th>
<th>Investment Sequence</th>
<th>Expected Net Present Value (x10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>f = 1.0</td>
<td>α₁, α₂, α₃</td>
<td>Λ₁, Λ₂, Λ₅, Λ₇</td>
<td>-1.700, 3.012, 8.988</td>
</tr>
<tr>
<td></td>
<td>Stochastic</td>
<td>Λ₂, Λ₄</td>
<td>1.872</td>
</tr>
<tr>
<td>f = 10.0</td>
<td>α₁, α₂, α₃</td>
<td>Λ₁, Λ₂, Λ₅, Λ₇</td>
<td>-1.752, 4.718, 8.678</td>
</tr>
<tr>
<td></td>
<td>Stochastic</td>
<td>Λ₁, *</td>
<td>2.239</td>
</tr>
</tbody>
</table>

* - indicates conditional alternatives
zonal accessibility. Accessibility is a direct function of the total number of trips (which are equal for each of the directions, i to j and j to i) but also inversely proportional to the zonal population. Thus, investment capacity will have a greater effect on the zone with the smaller population, and therefore induce a shift in population to that zone. Additionally, since demand is a function of the population product, a larger allocation to the smaller zone (of a constant total population increase) results in an increased flow volume between the two zones. Thus, benefits and costs at any stage depend not only on the demand level of the previous stage, but on the investment alternative as well.

The results shown in the table demonstrate the change in the investment policy for low, mean, and high demand spines; and more importantly, that for the full stochastic problem, the solution has changed from the non-sequential static uncertainty sequence of the previous section to a true sequential strategy approach. For $f = 1.0$, a modest accessibility factor, a mean demand state variable deterministic search produces alternative sequence $\Lambda_1^{e} - \Lambda_5^{e}$ (no expansion in stage 1 with investment in a new 2 lane in location G in stage 2) with net present value of $3.012 \times 10^6$. For the low and high state variables, we get proportionately less and more investment selected, respectively. Running a full stochastic tree search over all alternatives and all state variables produces the unconditional sequence $\Lambda_2^{e} - \Lambda_4^{e}$ (expansion in stage 1 of the existing 4 lane facility to an improved facility and further expansion in stage 2 of the improved facility to another improved level) with an expected present value of $1.872 \times 10^6$. Checking
the unconditional sequence $\Lambda_2 - \Lambda_4$ against a non-sequential static uncertainty forced sequence run, as described in Section 5.3.4, produces the same expected value.

Increasing the accessibility factor to $f = 10.0$ and running all state variables separately as deterministic tree searches produces the sequences $\Lambda_1 - \Lambda_2$, $\Lambda_1 - \Lambda_9$ and $\Lambda_5 - \Lambda_9$. A full stochastic tree search results in a conditional strategy $\Lambda_1^*$ (ENPV = $2.239 \times 10^6$). Running a non-sequential static uncertainty run over a number of possible sequences produces $\Lambda_1 - \Lambda_9$ with a net expected value of $0.8363 \times 10^6$ as the best investment sequence. Thus, the non-sequential solution is the same as the solution chosen for the mean state variable level for this example, but significantly different in terms of expected value ($0.8363 \times 10^6$ vs $4.718 \times 10^6$). The strategy approach predicts the same initial period investment but with the second stage depending on what demand level actually occurs at the end of the first stage, and with a substantial increase in expected value over the non-sequential solution.

Other values of $f$ produce increasingly higher and higher capacity improvements, because of the increase in volume due to a shifting population which induces more volume which induces more population, and so on.

In summary, for $f$ values above 2.5, a full stochastic tree search indicates that the best initial investment is the same as that predicted by a mean state variable deterministic search, but that the second period investment is dependent on what actually occurs in stage 1. Thus, for transport investments with long-run supply-demand dependencies, the investment problem becomes similar to the inventory control problem of the previous section and a true "tracking adaptive" solution is indicated. We would expect this effect to be more pronounced in the multi-zone network case where shifts in population and economic activity result not only in a different magnitude of volume, but in a different spatial allocation as well.
5.3.6 The Sensitivity of Investment Policies to Changed Initial Conditions

Before turning to the results of the approximating procedures developed to make application of the extensive form feasible for large problems, we pause first to consider the effects of varying one of the initial basic conditions of the example problem. Clearly, additional sensitivity testing is required to see how investment patterns change for a number of different conditions; for example, how are staging policies affected by the degree of scale economies present in the technology, how does the measure of benefits affect the choice (throughout we have used consumer surplus; it would be desirable to test a number of different measures that are commonly suggested in the literature such as minimum cost, maximum benefit-cost ratio, etc.), and so on. However, we leave this for future experiments with the model. In this section, we consider variations in investment due to changes in the price-elasticity parameter, \( \beta \).

Changing the parameter \( \beta \), which measures the responsiveness of demand to changes in the price of the trip, results in a changed investment pattern, and changes in expected value, as shown in Table 5-12. These results are again based on the two period model with stages of 5 and 10 years in length, a growth rate of 3% per year for population and an interest rate of 5%.

The results shown in the table represent a sensitivity analysis on \( \beta \) with values of 0.1, 1.0 and 2.0. The magnitude-of-demand parameter, \( \alpha_k \), still remains as the uncertain state variable.
Table 5-12
Testing the Sensitivity of Investment to Elasticity of Demand

<table>
<thead>
<tr>
<th>Demand Level</th>
<th>Inelastic 1.1</th>
<th>Unit Elastic 1.0</th>
<th>Elastic 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Investment Sequence Stage</td>
<td>ENPV (x10^6)</td>
<td>Investment Sequence Stage</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>( \Lambda_1 ), ( \Lambda_2 )</td>
<td>-1.412</td>
<td>( \Lambda_1 ), ( \Lambda_2 )</td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>( \Lambda_2 ), ( \Lambda_4 )</td>
<td>3.631</td>
<td>( \Lambda_5 ), ( \Lambda_5 )</td>
</tr>
<tr>
<td>( \alpha_3 )</td>
<td>( \Lambda_5 ), ( \Lambda_5 )</td>
<td>10.321</td>
<td>( \Lambda_5 ), ( \Lambda_7 )</td>
</tr>
<tr>
<td>Stochastic</td>
<td>( \Lambda_2 ), ( \Lambda_4 )</td>
<td>3.242</td>
<td>( \Lambda_2 ), ( \Lambda_4 )</td>
</tr>
</tbody>
</table>
Each $\beta_j$ therefore requires its own unique set of $\alpha_k$ factors in order to match the demand function with existing volumes.

Looking at each of the state variable levels for different $\beta$ values, we observe no significant change in the investment sequence for state variable $\alpha_1$ (sequence $\Lambda_1 - \Lambda_2$ in all cases), a significant change for state variable $\alpha_2$, where the investment sequence changes from $\Lambda_2 - \Lambda_4$ to $\Lambda_1 - \Lambda_9$ to $\Lambda_5 - \Lambda_5$ with a range of net present value of $1.667 \times 10^6$. Similarly for state variable $\alpha_3$ there is a difference both in the investment sequence and the net present value.

Running the full stochastic decision tree for the nonsequential static uncertainty for each $\beta$ level is also shown in Table 5-12. Going from .1 to 1.0 produces no change in the investment sequence ($\Lambda_2 - \Lambda_4$ in both cases) but, it does produce a drop in expected net present value of about 32%. This is simply due to changes in the loss associated with the sequence $\Lambda_2 - \Lambda_4$ for high and low demand $\alpha$ spines for different $\beta$'s. Interestingly enough, although the expected net present value is different from the net present value for the mean state variable level (as we would expect), the sequence of investment remains the same ($\Lambda_2 - \Lambda_4$ for $\beta = .1$ and $\Lambda_1 - \Lambda_9$ for $\beta = 2.0$) for the two extreme values. In these cases, the use of a mean state variable in place of a full tree search would give the correct investment sequence although it greatly overestimates the expected value. For $\beta = 1.0$ however, not only is there a difference in expected value between the non-sequential static uncertainty optimal sequence and the mean sequence ($2.209 \times 10^6$ to $2.721 \times 10^6$) but there is also a different sequence suggested for
investment (the suggested investment is improve the existing alterna-
tive to alternative $\Lambda_2$, then expand capacity in the second stage to
alternative $\Lambda_4$, another expansion of the existing facility). Both
the mean $\alpha$ level and the high $\alpha$ level suggest an initial investment
in alternative $\Lambda_5$ (this is the new 2 lane in location G).

From these initial runs, we conclude that the value of the
price-elasticity parameter can change both the decision and the
expected value produced by the model. The variations in $\beta$ tested
are somewhat extreme and we would expect that the elasticity value
is known to a much greater precision than this range. For example,
it may be elastic, unit elastic, or inelastic, and uncertainty would
range over one of these three areas but not all.

Extending capacity expansion time-staging problems to the
network problem, we expect the parameter $\beta$ to have a much greater
effect on the investment alternatives chosen. Recall that $\alpha$ essen-
tially controls the magnitude of demand—how much traffic we can
expect between i and j—and $\beta$ affects the distribution of traffic—
how much each j receives of the total volume emanating from a par-
ticular i. Large $\beta$ values tend to distribute more of the total
volumes (depending of course, on the network structure) to locations
close to the originating node i. Small $\beta$ values imply demand is
relatively insensitive to distance, or the price, of travel and,
therefore, volumes, in general, get distributed more evenly over a
larger area.

273
Using the $\alpha$ factor as the uncertain state variable in the network problem increases the magnitude of volumes which may now shift volumes from certain paths to others, but the question is still primarily one of how much capacity and when. If we also allow $\beta$ to be an uncertain state variable, then we also introduce, in a very large way, the spatial question, i.e. for small $\beta$, we may have a particular set of investments that are optimal (short links with high capacity); for high $\beta$, the optimal investment pattern may be a different set of investments which can better accommodate a longer set of trip patterns. Finally, using the full stochastic tree with $\beta$ as the uncertain state variable may produce an investment pattern with a mixed set of links, or it may be a set of links different from either those optimal for the high or low $\beta$ levels. Additionally, it would be desirable to explore the effects of multiple uncertain state variables (both $\alpha$ and $\beta$ uncertain), which may produce results still different again.
5.4 Experimental Results with the Proposed Approximation Procedures

The previous section has presented a series of runs designed to explore some of the major philosophical issues of investment planning under varying conditions of deterministic and stochastic demand assumptions. The example, although extensive in terms of the number of calculations when one considers the size of the tree generated for a 2 or 3 period model, is still rather simple in terms of number of state variable classes, stages, etc. However, we will continue using this example since we can then make comparisons between the policies based on the approximative model of Chapter 4.0 and the true optimal policies.

The procedures used in this section to reduce the amount of search were defined previously as pruning rules and terminal evaluation functions. Briefly, their purpose is first, to decide at what point in the tree the full solution space of future alternatives can be approximated and second, to produce an estimate of the utility of these future alternatives without actually performing the full set of calculations and comparisons. Note that most of these procedures can be used as overall macro-search procedures which implicitly prune this space before the search begins (for example, we can reduce the model to a 2 stage, 3 state variable tree even before beginning the search), or as explicit pruning rules and terminal functions within a given decision tree solution space.

In this section, we will first explore the difference in investment patterns for various terminal evaluation functions when used as
macro-search procedures and compared to the results of Section 5.3. We then will consider the effects of these functions when incorporated as terminal evaluation functions to be called automatically when pruning is indicated within a run.

5.4.1 Horizon Approximation

Using the stage aggregation technique as a horizon approximation procedure, we first evaluated the full 44 plans using a single period model of 15 years in length. After extensive exploration using all state variables, it became obvious that location H alternatives were clearly dominated by location G alternatives for both the 5% and 10% discount rates, using the 3 state variable model. We therefore eliminated it from further consideration in all of the following experiments. Although it is possible that we may have eliminated the best solution by ignoring the staging possibilities for H, it is consistent with the remaining approximating procedures to have made this approximation at this stage.

5.4.2 Results using the Myopic Search Rule

With the alternative space reduced from 44 plans to 24, we now turn to a comparison of specific approximating procedures with the global search rule. This section compares a myopic search procedure with the global search technique used in Section 5.3.

Reducing the search from a full global tree search, i.e., using the averaging-out and folding-back global procedure of Chapter 3.0, to a myopic, period by period maximization is nothing more than changing from a global optimization to a sub-optimization procedure.
In some cases, if the problem is separable, the rule will produce a solution that coincides with the global rule. Recall from Section 4.5 that the difference between a myopic procedure and a global one is that the former assumes only the current period results are of any importance in the decision at the current stage. Thus, all future decisions are ignored for the moment. The global rule, on the other hand, considers all possible future decisions before making a decision at the current stage.

There are two variants on the basic myopic rule incorporated in DECISION. The first determines the best decision using the full set of probability classes at the current stage, and then branches to the next stage using only the mean state variable level. The second version branches to the following stage using all state variables and proceeds to determine the best decision myopically for all state variables. Thus, the former rule reduces the space of solutions substantially more than the latter, but at a decreased degree of accuracy in measuring the expected value.

Using the first version of this myopic decision procedure, we compared the results obtained with the results of the global rule from the previous section for 2 and 3 stage models. These results are contained in Table 5-13.

The results of the 2 stage model differ from the global results for the mean and high state variable spines but are identical for the low demand spine. For example, for the state variable level \( \alpha_2 = 0.035 \times 10^{-6} \), the myopic and global rules give different initial
Table 5-13
Results Comparing the Myopic and Global Rules for 2 and 3 Stage Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Search Procedure</th>
<th>State Variable Level</th>
<th>Investment Sequence Stage</th>
<th>ENPV (x10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 stage (5 10)</td>
<td>Myopic</td>
<td>$\alpha_1 = 0.015 \times 10^{-6}$ $\alpha_2 = 0.035 \times 10^{-6}$ $\alpha_3 = 0.050 \times 10^{-6}$</td>
<td>$\Lambda_1 \quad \Lambda_2$ $\Lambda_2 \quad \Lambda_4$ $\Lambda_5 \quad \Lambda_7$</td>
<td>-1.594 2.672 9.991</td>
</tr>
<tr>
<td></td>
<td>Stochastic</td>
<td></td>
<td>$\Lambda_2 \quad \Lambda_4$</td>
<td>2.209</td>
</tr>
<tr>
<td>2 stage (5 10)</td>
<td>Global</td>
<td>$\alpha_1 = 0.015 \times 10^{-6}$ $\alpha_2 = 0.035 \times 10^{-6}$ $\alpha_3 = 0.050 \times 10^{-6}$</td>
<td>$\Lambda_1 \quad \Lambda_2$ $\Lambda_5 \quad \Lambda_5$ $\Lambda_5 \quad \Lambda_7$</td>
<td>-1.594 2.721 9.991</td>
</tr>
<tr>
<td></td>
<td>Stochastic</td>
<td></td>
<td>$\Lambda_2 \quad \Lambda_4$</td>
<td>2.209</td>
</tr>
<tr>
<td>3 stage (5 5 5)</td>
<td>Myopic</td>
<td>$\alpha_1 = 0.015 \times 10^{-6}$ $\alpha_2 = 0.035 \times 10^{-6}$ $\alpha_3 = 0.050 \times 10^{-6}$</td>
<td>$\Lambda_1 \quad \Lambda_1 \quad \Lambda_2$ $\Lambda_2 \quad \Lambda_3 \quad \Lambda_4$ $\Lambda_5 \quad \Lambda_5 \quad \Lambda_7$</td>
<td>-1.655 2.650 9.875</td>
</tr>
<tr>
<td></td>
<td>Stochastic</td>
<td></td>
<td>$\Lambda_2 \quad \Lambda_4 \quad \Lambda_4$</td>
<td>2.124</td>
</tr>
<tr>
<td>3 stage (5 5 5)</td>
<td>Global</td>
<td>$\alpha_1 = 0.015 \times 10^{-6}$ $\alpha_2 = 0.035 \times 10^{-6}$ $\alpha_3 = 0.050 \times 10^{-6}$</td>
<td>$\Lambda_1 \quad \Lambda_1 \quad \Lambda_2$ $\Lambda_5 \quad \Lambda_5 \quad \Lambda_5$ $\Lambda_9 \quad \Lambda_9 \quad \Lambda_9$</td>
<td>-1.604 2.782 10.07</td>
</tr>
<tr>
<td></td>
<td>Stochastic</td>
<td></td>
<td>$\Lambda_2 \quad \Lambda_3 \quad \Lambda_4$</td>
<td>2.183</td>
</tr>
</tbody>
</table>

NOTE: ENPV - Expected Net Present Value
period investments (Plan $A_2$ versus $A_5$) but with a difference of only $0.049 \times 10^6$ in total net benefits. For the overall stochastic tree search, the alternative selected is identical for both rules with no difference in expected value.

Turning to the 3 stage results of the model in the bottom of the table, with the same horizon of 15 years, we observe much the same comparisons except that now the myopic rule applied to the mean demand level has chosen the same sequence as the global for the full stochastic tree search.

In this set of experiments, we have used the myopic rule as an overall search procedure (i.e. a macro-search procedure) by using the routine MYOPDN from the very beginning of the tree search. Using this rule as a macro-search procedure in larger examples, we would expect that it would produce quite different results from the global results in the majority of the cases. However, the myopic rule was not originally intended to operate solely as a macro-search procedure for our capacity expansion model.\(^1\) It was designed to act as a terminal evaluation function, which, at a point in the tree determined by the set of pruning rules, will be invoked to estimate the value of the future from this point on, without continuing with the full global search. As we will show in Section 5.5, the myopic rule cuts down dramatically on the amount of computation time, and from preliminary indications, still gives a good estimate of the expected

\(^1\)Marglin [20] has demonstrated the losses that can occur by using this rule in such a manner.
value when used only for points further out in the tree. For such points, we suspect the error in approximating the optimal solution by using a myopic rule will have little effect on the initial decision, especially if a strategic approach is taken.

One possible extension to this rule is to use the myopic procedure as a k-stage global rule. In other words, we perform a global search over the next k stages ignoring the future from stage k on. Thus we are actually looking myopically at a k-stage model. Additionally, it could be used in conjunction with the stage aggregation function of the following section.

5.4.3 Results of Stage Aggregation

Again, as in the case of the myopic search rule of the previous section, this rule may be employed as either a macro-search procedure to be used from the very beginning of a tree search or as a terminal evaluation function to be called automatically when pruning is indicated by the set of pruning rules. The purpose of this procedure is to reduce the space of alternatives by limiting the number of possible investment periods. The results from a series of experiments using this technique are shown in Table 5-14.

Comparing a 3 stage model with stages of 5 years in length to a 2 stage model of 5 and 10 years we observe no significant difference in total benefits for any of the state variable levels, and only a change in one instance of the investment sequence from a full 4 lane investment, A to a staged investment sequence, A -A.
<table>
<thead>
<tr>
<th>Model Type</th>
<th>Demand Levels</th>
<th>Investment Sequence</th>
<th>ENPV Benefits (x10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \alpha_1 )</td>
<td>Stage 1 Stage 2 Stage 3</td>
<td>-1.598</td>
</tr>
<tr>
<td>2 stage</td>
<td>( \alpha_2 )</td>
<td>( \Lambda_1 ) ( \Lambda_2 )</td>
<td>2.721</td>
</tr>
<tr>
<td></td>
<td>( \alpha_3 )</td>
<td>( \Lambda_5 ) ( \Lambda_5 )</td>
<td>9.998</td>
</tr>
<tr>
<td>3 stage</td>
<td>( \alpha_1 )</td>
<td>( \Lambda_1 ) ( \Lambda_1 ) ( \Lambda_2 )</td>
<td>-1.632</td>
</tr>
<tr>
<td></td>
<td>( \alpha_2 )</td>
<td>( \Lambda_5 ) ( \Lambda_5 ) ( \Lambda_5 )</td>
<td>2.872</td>
</tr>
<tr>
<td></td>
<td>( \alpha_3 )</td>
<td>( \Lambda_9 ) ( \Lambda_9 ) ( \Lambda_9 )</td>
<td>10.099</td>
</tr>
</tbody>
</table>

ENPV - expected net present value
5.4.4 Effects of Variations in the Number of Classes and Selected Probability Traces

To give an indication of the effects of the number of probability classes on the investment decision, we next compare a series of experiments using the same general probability distribution shape, but with variations in the number of discrete class approximations to this distribution. Table 5-15 contains the results for 1, 3 and 5 classes with the state variable level ranging from $0.015 \times 10^{-6}$ to $0.050 \times 10^{-6}$.

Comparing the stochastic results with the optimal sequence for the mean state variable shows a different sequence in all three cases, and not surprisingly, a substantial difference in the expected net present value from the net present value for the mean state variable.

From Section 4.5, we recall that a state variable spine is defined to be a selected sequence of multi-period state variables of constant magnitude. By selecting a mean state variable spine, we see from the previous table that in some cases we could produce a relatively good estimate of the utility of any point in the tree, without exploring all possible combinations.

A comparison of the full tree exploration shown in the above table with the mean demand level demonstrates the difference between the deterministic and stochastic assumptions. For the example shown here, the results for the mean demand spine significantly overestimate the investment required. It overestimates the total net benefit by
Table 5-15
Effects of Variations in Number of Probability Classes

<table>
<thead>
<tr>
<th>No. of Class Elements</th>
<th>State Variable Levels (x10^{-6})</th>
<th>Investment Sequence</th>
<th>Deterministic NPV (x10^6)</th>
<th>Stochastic ENPV (x10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>sequence</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.035</td>
<td>( \Lambda_5 \Lambda_5 )</td>
<td>2.721</td>
<td>N.A.</td>
</tr>
<tr>
<td>3</td>
<td>.015</td>
<td>( \Lambda_1 \Lambda_2 )</td>
<td>-1.598</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \Lambda_5 \Lambda_5 )</td>
<td>2.721</td>
<td>( \Lambda_2 \Lambda_4 )</td>
</tr>
<tr>
<td></td>
<td>.050</td>
<td>( \Lambda_5 \Lambda_7 )</td>
<td>9.998</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.015</td>
<td>( \Lambda_1 \Lambda_2 )</td>
<td>-1.598</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.025</td>
<td>( \Lambda_2 \Lambda_3 )</td>
<td>- .5744</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.035</td>
<td>( \Lambda_5 \Lambda_5 )</td>
<td>2.721</td>
<td>( \Lambda_2 \Lambda_4 )</td>
</tr>
<tr>
<td></td>
<td>.042</td>
<td>( \Lambda_5 \Lambda_7 )</td>
<td>1.280</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.050</td>
<td>( \Lambda_5 \Lambda_7 )</td>
<td>9.998</td>
<td></td>
</tr>
</tbody>
</table>

Notes:

N.A. - not applicable
NPV - net present value
ENPV - expected net present value
$.257 \times 10^6$ for the 3 state variable model and $.571 \times 10^6$ for the 3 state variable model and $.571 \times 10^6$ for the 5 state variable version. A different initial investment is selected in both cases. However, as an explicit terminal evaluation function this procedure can in some instances be a useful way of collapsing parts of the full tree or as a partial initial analysis. In this case, a deterministic problem replaces this part of the tree and provides a good, approximate estimate for the expected value from this point on.

5.4.5 Comparison of Myopic and Stage Aggregation as Terminal Evaluation Functions

In this final series of experiments, we now turn to the use of the terminal evaluation functions as automatic procedures to be used when one of the pruning rules indicates a portion of the tree will offer little in the way of changing the expected value from this point on.

In order to be able to explore a fairly extensive tree in terms of number of stages, we reduce the 14 plans of the previous sections to only 4. However, we will select the projects to be representative of the full range of capacities; these are the existing alternative (1), and the largest capacity improvement to that alternative (4), the new 2 lane facility (5), its expansion to 4 lanes (7), and the full 4 lane immediate expansion facility (9). We further constrain the problem to a single investment sequence using the state transformation matrix, i.e. the invest-
ment plans consist of the mutually exclusive set \( A_1, A_4, A_5-A_7, \) and \( A_9. \) The horizon, \( \hat{N}, \) is 20 years with 4 stages of 5 years in length each. The state variable distribution assumed to be is \( p_k(\phi_{kt}) = [.2 .5 .3]. \)

We now introduce the pruning probability sets I, II, II, and IV, in a series of experiments to compare the approximate solutions, produced by applying the pruning rules and terminal functions, with the full tree search. These probabilities are shown in Table 5-16.

### Table 5-16

<table>
<thead>
<tr>
<th>Stage</th>
<th>Minimum Pruning Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
</tr>
<tr>
<td>1</td>
<td>0.55</td>
</tr>
<tr>
<td>2</td>
<td>0.65</td>
</tr>
<tr>
<td>3</td>
<td>0.75</td>
</tr>
<tr>
<td>4</td>
<td>0.85</td>
</tr>
</tbody>
</table>

The full stochastic tree global solution to this problem (produced without the pruning rules) is shown in Table 5-17. The optimal sequence for the mean demand spine is \( A_5-A_5-A_5-A_7 \) but changes to \( A_1-A_1-A_9-A_9 \) for the full stochastic tree solution.
Table 5-17

Global Solution for Comparison with the Myopic Search Rule and Stage Aggregation

<table>
<thead>
<tr>
<th>Demand Level (x10^{-6})</th>
<th>Investment Sequence Stage</th>
<th>Expected Net Present Value (x10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>$\alpha_1 = 0.015$</td>
<td>$\Lambda_1$</td>
<td>$\Lambda_1$</td>
</tr>
<tr>
<td>$\alpha_2 = 0.035$</td>
<td>$\Lambda_5$</td>
<td>$\Lambda_5$</td>
</tr>
<tr>
<td>$\alpha_3 = 0.050$</td>
<td>$\Lambda_9$</td>
<td>$\Lambda_9$</td>
</tr>
<tr>
<td>Stochastic</td>
<td>$\Lambda_1$</td>
<td>$\Lambda_1$</td>
</tr>
</tbody>
</table>
Proceeding from set I through IV implies that we prune less and less of the tree i.e. we are exploring increasingly larger and larger trees. For example, set I will prune all of the tree at the end of the first stage. Set II (and III) prunes at different stages depending on the probability at that stage.

The two terminal functions we use to estimate the future from the pruning points on are the myopic rule and stage aggregation. The results of applying these two terminal functions and pruning rule set I are contained in Table 5-18.

Note that the solution for the full stochastic tree search is not too different for stage aggregation but quite different for the myopic search used from the end of stage 1 to the horizon. The global solution first period investment is $\Lambda_1$ with a total expected value of $5.607 \times 10^6$; the stage aggregated solution for pruning probability set I is also $\Lambda_1$ with expected value equal to $6.188 \times 10^6$; however, the myopic solution is $\Lambda_4$ with an expected value of only $2.800 \times 10^6$. Notably, computation time has been cut from 65 seconds for the global to 7.5 seconds for both the stage aggregation and myopic terminal functions.

Performing the same experiments for the second pruning set (II), we see from Table 5-19 that pruning now occurs at different stages depending on the history, i.e., either the path product probability to that point, or the stage probability, as we described in Chapter 4.0. For example, on the $\alpha_1$ spine, stage aggregation occurs at the
Table 5-18

Results with Pruning Rule Set I
and the State Aggregation and Myopic Terminal Functions

<table>
<thead>
<tr>
<th>Terminal Evaluation Function</th>
<th>Demand Level ( (x10^{-6}) )</th>
<th>Investment Sequence</th>
<th>Expected Value ( (x10^6) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage Aggregation</td>
<td>( \alpha_1 = 0.015 )</td>
<td>( \Lambda_1 ) - - ( \Lambda_1 )</td>
<td>- 1.754</td>
</tr>
<tr>
<td></td>
<td>( \alpha_2 = 0.035 )</td>
<td>( \Lambda_1 ) - - ( \Lambda_9 )</td>
<td>5.946</td>
</tr>
<tr>
<td></td>
<td>( \alpha_3 = 0.050 )</td>
<td>( \Lambda_5 ) - - ( \Lambda_7 )</td>
<td>15.190</td>
</tr>
<tr>
<td></td>
<td>Stochastic ( \Lambda_1 ) - - ( \Lambda_9 )</td>
<td></td>
<td>6.188</td>
</tr>
<tr>
<td>Myopic Search</td>
<td>( \alpha_1 = 0.015 )</td>
<td>( \Lambda_1 ) ( \Lambda_1 ) ( \Lambda_1 ) ( \Lambda_1 )</td>
<td>- 1.724</td>
</tr>
<tr>
<td></td>
<td>( \alpha_2 = 0.035 )</td>
<td>( \Lambda_4 ) ( \Lambda_4 ) ( \Lambda_4 ) ( \Lambda_4 )</td>
<td>4.286</td>
</tr>
<tr>
<td></td>
<td>( \alpha_3 = 0.050 )</td>
<td>( \Lambda_5 ) ( \Lambda_5 ) ( \Lambda_7 ) ( \Lambda_7 )</td>
<td>14.980</td>
</tr>
<tr>
<td></td>
<td>Stochastic ( \Lambda_4 ) ( \Lambda_4 ) ( \Lambda_4 ) ( \Lambda_4 )</td>
<td></td>
<td>2.800</td>
</tr>
</tbody>
</table>

- = stage aggregated

288
end of the first stage; on the $q_3$ spine, it occurs at the end of the second; and for the mean spine it only occurs at the end of the 3rd stage (and the solution produced is identical to the global solution for this spine).

Table 5-19
Results of Stage Aggregation using Pruning
Set II

<table>
<thead>
<tr>
<th>Demand Level $(x10^{-6})$</th>
<th>Investment Sequence Stage 1 2 3 4</th>
<th>Expected Net Present Value $(x10^5)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>$\Lambda_1$ - - $\Lambda_1$</td>
<td>1.754</td>
</tr>
<tr>
<td>$a_2$</td>
<td>$\Lambda_5$ $\Lambda_5$ $\Lambda_5$ $\Lambda_7$</td>
<td>5.394</td>
</tr>
<tr>
<td>$a_3$</td>
<td>$\Lambda_9$ $\Lambda_9$ - $\Lambda_9$</td>
<td>15.080</td>
</tr>
<tr>
<td>Stochastic</td>
<td>$\Lambda_1$ * * *</td>
<td>5.921</td>
</tr>
</tbody>
</table>

Continuing with the experiments for pruning rule sets II through IV, we see a significant difference in the two terminal functions depending on where they are applied in the tree. We summarize the results for the stochastic tree searches for these experiments in Table 5-20.
Table 5-20
First Period Investments and Total Expected Values
for Pruning Sets I through IV

<table>
<thead>
<tr>
<th>Pruning Set</th>
<th>Stage Aggregation</th>
<th>Myopic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First Stage</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Investment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Expected Value</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(x10^6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Time (Seconds)</td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------</td>
<td>--------</td>
</tr>
<tr>
<td>I</td>
<td>Λ_1</td>
<td>6.188</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Λ_4</td>
<td>2.800</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>II</td>
<td>Λ_1</td>
<td>5.921</td>
</tr>
<tr>
<td></td>
<td></td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Λ_1</td>
<td>5.352</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>III</td>
<td>Λ_1</td>
<td>5.921</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Λ_1</td>
<td>5.607</td>
</tr>
<tr>
<td></td>
<td></td>
<td>45</td>
</tr>
<tr>
<td>IV</td>
<td>Λ_1</td>
<td>5.607</td>
</tr>
<tr>
<td></td>
<td></td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>Λ_1</td>
<td>5.607</td>
</tr>
<tr>
<td></td>
<td></td>
<td>65</td>
</tr>
</tbody>
</table>

From the results of these experiments, we see that both terminal functions provide reasonable estimates of the expected value, and also choose the same initial investment, for the second pruning set of (.30 , .40 , .50 , .60). However, using the pruning rule set which essentially prunes at the first stage, there is a significant difference between the global solution, both in expected value and in the selected alternative, for the myopic rule, and in expected value for the stage aggregation terminal function. Using a pruning rule set which prunes less and less of the tree, gives progressively better solutions (they approach the global) but at increasingly higher costs of computation.

290
Comparing the myopic and global rules, it appears that the myopic is almost as good as stage aggregation beyond the second pruning rule set and reduces the computation time even more significantly than stage aggregation. Stage aggregation of time periods would appear to be most useful in cases where there are no dramatic increases in demand or shifts in flow patterns. In other words, if the response surface over time is relatively flat, we can expect little to be lost by skipping some intermediate periods for investment.

Although we have used the pruning probability themselves as rules to tell us where to cut the tree, we might also usefully consider other alternative rules; for example, in place of the stage and product probability rules, we might incorporate both the return and the probability at this stage. A low probability event may produce consequences that are in a sense catastrophic - we really should be concerned with the contribution to expected value and not just the probability. In the present version of the pruning rule subroutine, we are actually using the probability alone as a surrogate for this approach. A number of additional combinations of rules (a k-level myopic as a look-ahead heuristic, exploring maximum probability spines, and so on) might also be explored for their usefulness in particular situations.

Clearly, the results of these experiments have indicated that the full tree search can be approximated quite satisfactorily using a variety of different terminal evaluation functions. More experimentation is needed, however, to indicate which rules and functions are most appropriate for specific parts of the tree (and for different problem types.)
5.5 Comparison of Computation Times for Variations in Problem Size

As we indicated in Section 5.3, computation time proved to be a severe constraint in adequately testing the use of the model for questions of investment issues under uncertainty. Because of the time-consuming nature of the network flow equilibrium determination, activity shift computations, etc., and the cost of this time in terms of computer expenses, we had to restrict the problem size to something less than we would have liked. We first limited the number of stages to 3, then to 2, and the number of alternatives from 44 to 24 for the majority of runs. The reason for this is obvious when we consider the computation times for variations in the number of state variables, alternatives and stages.

On the one hand, the size of the tree and the amount of time required is a good indication that heuristics for pruning parts of the tree are absolutely essential if such an approach is to be used in investment planning. On the other hand, because of the cost of running stochastic tree searches, it is difficult to compare the effectiveness of these rules against the optimal solution in an absolute sense. For a realistic problem size—anything more than 3 state variables, 2 or 3 stages and 24 plans—the amount of time required to determine the optimal solution is prohibitive.

Even with this constraint, however, our computational results have clearly borne out what we expected a priori, i.e., uncertainty in demand can have an effect on both the actual investment sequence chosen and its expected value. They have also suggested some
interesting conclusions, some expected and some which we did not expect. The purpose of this section is to report on the computation times that were observed for the previous two sets of experiments.

The programs representing the decision tree and simulation models were run on an IBM 360/67 time-sharing system. The results of these experiments can only be reported to be approximate estimates since actual times depend on the amount of output requested and even the time of day when experiments were conducted. During the day, there is a greater number of users on the system which affects how much shuffling in and out of core programs must do, and which counts in the reported times. For the most part, we have attempted to minimize this by reporting on times when there was a minimal load on the system or by adjusting the times reported for consistency.

The execution time for the experiments of Section 5.3 are summarized in Figure 5-10 for both deterministic and stochastic tree searches. Results for trees with 5, 24, and 44 plans and 1, 3, and 5 state variable classes are reported.

The results are what one would expect based on the calculations of tree size of Section 3.2.7. For anything beyond 4 or 5 stages, the time required for a deterministic run for 24 plans begins to become prohibitive for extensive testing; for example, a 6 period model requires 3.5 minutes. For the stochastic version, the

\[\text{At $325.00/hour, this is approximately $20.00 for a full global search.}\]
Legend:

- $A_1$ - 5 Plans, Deterministic
- $A_3$ - 5 Plans, 3 State Variables
- $B_1$ - 24 Plans, Deterministic
- $B_3$ - 24 Plans, 3 State Variables
- $B_5$ - 24 Plans, 5 State Variables
- $C_1$ - 44 Plans, Deterministic

Figure 5-10
Computation Time for Various Decision Tree Sizes
same 3.5 minutes constrains us to a 3 period model.

Increasing the number of plans to 44 by including location H and its alternatives increases the computation time drastically. For a deterministic run, a 3 period model had to be cut off after 7 minutes with still no solution. Similarly, increasing the number of state variable classes from 3 to 5 for a two period, 24 plan model increases the time from 21 to 55 seconds.

Results for the pruning rule and terminal evaluation experiments of Section 5.4 indicate quite substantial drops in these run times as we would expect. For a 3 period, 24 plan model, the myopic rule requires only an average of 6 seconds for a deterministic run (versus 15 seconds for a global search) and an average of 10 seconds for a stochastic search (versus 35 seconds for a global). Stage aggregation which reduces the 3 stage model to a 2 stage also reduces the search time from 15 to 8 seconds for a deterministic run and from 35 to 21 for the stochastic version.

Increasing the alternative plans from 24 to 44 showed little increase in run times for both myopic and stage aggregation compared to a 10-fold increase for the global rule.

5.6 Summary and Conclusions

This chapter has explored two basically different aspects of the transport investment problem using the project design case as an example: the first was concerned with the effects of uncertainty on investment decisions in terms of theoretical issues. Thus, we ignored
the computational problems associated with the combinatorial nature of
the multistage investment problem by operating with a rather limited
model. In the second series of experiments, we conducted some
additional experiments to explore the effects of a series of approxi-
mating procedures, termed pruning rules and terminal evaluation func-
tions, to be used when the combinatorial nature of the problem dominates.
The intent of this section is to briefly summarize this work and our
conclusions based on the experiments to date.

In terms of the first series of experiments concerned with
investment issues in principle, we first conducted a very simple
set of experiments on a restricted problem to look at what have
now become fairly familiar concepts in investment planning, i.e.,
absolute vs. comparative advantage (the pure timing problem), and
staging vs. scale economy tradeoffs. We then turned to the intro-
duction of uncertainty and showed the value of technology adaptation
for the single link case. The value of providing an adaptable
technology was shown to be that it broadens the range where adapta-
tion is desirable and it lowers the costs (or increases the benefits)
in the zone where expansion would occur in any event.

In the next series of experiments, we then turned to the
primary area of interest of this study - the sequential decision
model and investment planning policies under deterministic, non-
sequential stochastic and sequential decision problems. From the
results of this set of experiments, we conclude that there can be
significant differences for investment patterns between the deterministic and non-sequential stochastic models. Unlike the inventory control models, however, we also showed that there is no difference between the sequential and non-sequential versions of the transport investment problem when either there is no dependence between investment and the socio-economic growth of the region, or the state variable probability distribution is time-independent and stationary.

When a long run supply-demand dependence is introduced, however, we observed a tracking adaptive solution which, in the true sequential decision theory fashion, predicted a conditional decision for the second stage dependent on what actually occurs at the end of the first stage.

A final series of experiments indicated quite substantial changes in investment policies depending on the initial conditions. Differences in the price elasticity parameters was shown to greatly affect the time-staged investment pattern. These effects were shown to be handled in a straightforward manner by the model.

We then turned next to the computational problems of the investment problem under uncertainty. Using the same example of the previous section, we conducted some initial experiments designed to test the effects of specific heuristics designed to reduce the solution space.

Briefly, our conclusions based on these results have been encouraging from the point of view of large combinatorial problems.
For the project design case, the selected procedures presented have shown that they all are useful in reducing the solution space without significantly affecting the initial decision. We are reluctant to draw other more specific conclusions as to their value until more extensive testing is done with larger and more extensive network problems than have been tested to date. A number of procedures, such as linear and dynamic programming, branch and bound, etc. should prove useful as partial search procedures for the time-staging problem.

In summary, the conclusions we can reach from the previously described set of experiments are limited first, by the specific nature of the example chosen to demonstrate the sequential procedure and second, by the underlying nature of the project design problem itself for the transport case. The case of project design is generally much simpler than those found in water resources where capacities can take on an almost infinite number of scales and decisions to invest in one scale facility have a profound effect on future expansions (see Russel [21]). We did expect that there would be more variation in results due to the aggregate capacity or network effects for the transport example, Nonetheless, the single link example did provide us with a mechanism for testing the properties of the search procedure and acted as a vehicle for testing the programs. In the following chapter, we extend the basic model for the single case to include the effects of a learning adaptive model on the sequential decision procedure.
Chapter 6

AN ADAPTIVE SEQUENTIAL MODEL FOR CAPACITY EXPANSION UNDER UNCERTAINTY
Chapter 6
AN ADAPTIVE SEQUENTIAL MODEL FOR CAPACITY EXPANSION UNDER UNCERTAINTY

To this point, the structure of the sequential model has assumed full knowledge of the underlying randomness of the state variable by specifying a constant probability density function for the state variable, independent of the stage we are currently at, and of previous historical results. In most cases, however, the true underlying distribution will not be known, although it may be constant and independent from period to period. This chapter relaxes the requirement of known probability functions and develops an adaptive mechanism that allows the process to learn about the underlying structure of randomness through the technique of a Bayesian learning model. The purpose of this section will be to define this procedure, and to explore its implications on investment issues; specifically, how will the initial investment change when such an adaptive procedure is incorporated.

6.1 Restatement of the Demand Uncertainty

Before proceeding to extend the analysis to the case of unknown distributions, it will be useful to review the general sequential decision model with known distributions. Section 2.3 initially defined the single state variable, multistage decision structure as a sequence of uncertain state variables, one for each stage or time period with each specific value at a stage having a probability of occurrence. The state variable, $\phi$, was defined as some parameter or vector of parameters,
\[ \phi = f(\alpha, \beta) \text{ of a demand function of the form} \]

\[ v_{ij}^m = \alpha_0 (P_i^j P_j^i)^{\alpha_1} (Y_i^j Y_j^i)^{\alpha_2} (E_j^i E_i^j)^{\alpha_3} (c_{ij})^\beta \]

where \( v_{ij}^m \) = the demand for travel between nodes i and j by mode m

\( P_i, P_m \) = population at i and j at any point in time

\( Y_i \) = income at i at any point in time

\( E_j \) = employment at j at any point in time

\( c_{ij} \) = the price variable or cost of travel from i to j

\( \alpha, \beta \) = the demand function parameters

The state variable, \( \phi_k \), has an associated probability density function, \( p_k(\phi_k) \), which is constant from period to period and assumed to be known exactly. Therefore, although the flow volume will change as a function of the zonal socio-economic characteristics which change with time, and is therefore related to previous volumes through population and income growth, the value that the state variable, \( \phi_k \), takes on for any stage is independent of its value at a previous stage. Thus, the underlying stochastic structural relationship is assumed to be time-independent and stationary.

The above case is valid for certain kinds of processes. They must be well known and observable over a large number of trials. In many instances, however, even though we may assume the probability distribution is constant and time-independent, we may not have enough information about the distribution to state its form exactly.
Therefore using a "best guess" distribution with mean $\mu$, and variance $\sigma$, based on an initial calibration is acceptable for the initial stage of the process only. In the second stage, we would expect that the first period results, i.e. observing either a high or low volume different from the mean, would have some effect on our second stage estimate of the true distribution. If demand were low for 2 or 3 stages in a row, we would, in fact, expect to have a significantly different estimate of the underlying distribution at the 3rd stage.

In the real world, decisions are conditional on results of earlier periods, and the estimate of the underlying stochastic structure is revised as time unfolds. In the model so far, we have no mechanism for accounting for that revision in the sequential decision tree expected value calculations. If we know a priori that our distribution will be revised in real time as demands are observed, then the decision tree computations should account for that expected revision in its initial calculations.¹ This is accomplished using the initial estimate of the distribution from the previous stage and a conditional distribution of expected change given this previous result.

We have employed such a learning mechanism in the program DECISM in order that we may explore the consequences of such a revision process. The following section discusses this learning mechanism in principle and as we have applied it in DECISM.

¹In Bayesian decision theory terms, this is generally known as pre-posterior analysis.
6.2 Development of a Learning Adaptive Model for Capacity Expansion under Uncertainty

6.2.1 Introduction

The basis for the learning scheme discussed in the previous section is statistical decision theory, or more specifically, a subset of this general theory referred to as Bayesian decision theory (BDT)\(^1\) [1,2,3,4]. Both are subsumed within the more general sequential decision theory we have discussed in previous sections.

The Bayesian model is generally defined as consisting of experiment, result, action, state of nature \((e/z/a/\phi)\) for each stage of the process. Briefly, the model uses observations about the real world, \(z\), obtained from experiments, \(e\), to infer knowledge about a true state of nature variable, \(\phi\). This information can then be used to determine whether to collect more information to further improve our knowledge, or to take some immediate action, \(a\). Both have consequences associated with them that we should like to minimize or maximize. Using Bayes theorem, a revised estimate of the probability of a state of nature variable \((\phi)\) taking on a specific value can be computed using the previous estimate, \(p'(\phi)\), and a conditional distribution, \(g(z|\phi)\), which gives the probability of observing a result given the true value of the state of nature variable. The form of this revision process is given by Bayes theorem as

\(^1\)It is assumed that the reader (a) has an understanding of the basic mechanics of the traditional Bayesian framework, and (b) accepts the philosophy of the use of subjective probability distributions. For an introductory discussion, see Raiffa [3] or Hadley [4]. For a more advanced general discussion of applied adaptive theory, see Murphy [5] or Sawargi et al [6].
\[ p''(\phi | z, e) = \frac{p'(\phi) \cdot g(z | \phi, e)}{n(z)} \]

where

\[ p''(\phi | z, e) = \text{the revised or posterior distribution on the true value of the state of nature variable given an observation } z, \text{ and experiment, } e \]

\[ p'(\phi) = \text{the initial estimated distribution; the prior} \]

\[ g(z | \phi, e) = \text{a conditional distribution of observing result } z \text{ given the true value is } \phi \]

\[ n(z) = \text{a normalizing function to ensure the posterior is a true probability distribution, i.e. it sums to 1.} \]

The elements of this general problem can be structured in decision tree form as in Figure 6-1. The computational procedure is similar to the algorithm presented in Chapter 3.0. It first computes revised estimates of the state variable probability for all possible paths through the tree. It then uses the averaging-out and folding-back technique described previously to evaluate all branches of tree, until arriving back at the initial stage where all paths will then have been compared and a decision can be made; either collect more information, \( e_1 \), before deciding on what action to take, or take immediate action (proceed along the null experiment branch, \( e_0 \)).

Multiple stages or sequential experiments are evaluated by extending the tree and the calculations in a sequential fashion. In the case of transport planning, experiments correspond to information collection procedures, results to volume counts on links, actions to investment alternatives, and states of nature to the actual volumes
Figure 6-1
The Basic Bayesian Decision Tree
that occur.\footnote{See Johnson [7] for an initial exploratory analysis of a single period Bayesian model applied to link counting activities. See Khan [8] for the use of this approach for a 2-stage mass transit corridor expansion program analysis.}

Our decision model differs from the normal Bayesian decision tree in a number of respects. First, we are not so much concerned with information collection experiments per se as we are with multiple period sequential investment decisions. Therefore our tree has only two elements for every stage: an action - state of nature couple, which in our case represents an investment alternative, and a demand state. Second, since investments are durable, they can heavily influence the costs and benefits of future stages. Actions of future stages must therefore include the actions taken at previous stages.

Given this brief explanation of the difference between the multistage sequential decision tree of this study and the traditional (e/z/a/\phi) tree, we may now turn to a more formal mathematical notation for the process.

6.2.2 Mathematical Formulation of the Bayesian Approach for the Multistage Sequential Investment Problem

For any stage in the tree, we can identify a sequence of events that have occurred up to the \( t^{th} \) stage by the vector

\[
E_t = [e_1, e_2, \ldots, e_{t-1}]
\]

At any given stage, the event \( e_t \) is represented by the demand state variable vector \( \phi_t = [\phi_{1t}, \phi_{2t}, \ldots, \phi_{mt}] \) and can take on any one
of the specific $\phi_{kt}$ values with probability $p_k(\phi_{kt})$. Since we assume the process is completely random and independent of the stage index, $\phi_t$ is the same for all $t$, and the distribution over the events at a stage $t$ is given by:

$$P(e_t) = \{\Pr(e_t = \phi_k); k \in m\}$$

for any $t=1,2,...,T$.

With this notation and our assumption that state variables or events are independent of the stage index, we can now compute the probability of a particular event sequence occurring over time as simply the product of the probabilities of each event occurring separately or

$$P(e_1, e_2, \ldots, e_T) = \Pr(e_1 = \phi_1) \cdot \Pr(e_2 = \phi_j) \cdot \ldots \cdot \Pr(e_T = \phi_k)$$

This is essentially the technique used in Chapter 5.0 to compute the probability of any state variable sequence occurring.

The problem with using this joint probability for evaluating investment sequences is, as we stated previously, that it assumes the initial prior distribution over the state variable $p_k(\phi_k)$ is known exactly. If it is not known exactly, we would expect revision to occur at the end of the first stage. In order to account for this expected revision, we will incorporate the Bayesian learning scheme to produce a revised posterior for the second and subsequent stages based on the previous stage results.

This revised subjective probability distribution, using the conditional mechanism of BDT, is not, however, independent of the stage index. In other words, the true distribution is simply the product of
independent probabilities, but since we do not know the true distribution and are employing Bayes Theorem to improve our estimate of that true distribution, the joint subjective probability of any sequence of events or state variables occurring is not independent of the stage index. In fact, it is given by the product of the conditionals or revised distribution estimates as:

\[
\hat{Pr}(e_1, e_2, \ldots, e_T) = \hat{Pr}(e_1 = \phi_1) \cdot \hat{Pr}(e_2 = \phi_j | e_1) \cdot \ldots \cdot \hat{Pr}(e_T = \phi_k | e_1, e_2, \ldots, e_{T-1})
\]

For ease of notation, let us now define the conditional function:

\[
\hat{Pr}(e_t | e_1, e_2, \ldots, e_{t-1}) = \hat{Pr}(e_t = \phi_k | e_1, e_2, \ldots, e_{t-1}; k \in m)
\]

for any \(t=1, \ldots, T\)
as our present estimate for an event at the \(t^{\text{th}}\) stage. That is, in general, the revised conditional subjective distribution \(\hat{Pr}(e_t | e_1, e_2, \ldots, e_{t-1})\) of an event occurring at stage \(t\) will be different from the initial independent-of-stage distribution

\[
\hat{Pr}(e_t) = \hat{Pr}(e_t = \phi_k) = \hat{Pr}(\phi_k; k \in m)
\]

we originally assumed, and more closely approximates the true underlying distribution.

Using the above conditionals and the fact that the joint distribution is the product of two marginals, we can compute the subjective conditional probability of the \(t^{\text{th}}\) event, given the \((t-1)^{\text{st}}\) event as

\[
\hat{Pr}(e_t | e_{t-1}) = \frac{\hat{Pr}(e_t, e_{t-1})}{\hat{Pr}(e_{t-1})}
\]

308
The numerator $\hat{P}(e_t, e_{t-1})$ is given by summing over all possible sequences or

$$\hat{P}(e_t, e_{t-1}) = \sum_{\{E_1, E_2 \ldots E_{t-2}\}} \hat{P}(e_1, e_2 \ldots e_{t-1}, e_t)$$

The denominator is simply the unconditional of event $e_{t-1}$, given by computing

$$\hat{P}(e_{t-1}) = \sum_{\{E_1, E_2 \ldots E_{t-2}\}} \hat{P}(e_1, e_2 \ldots e_{t-2}, e_{t-1})$$

For computational purposes, it is sufficient to define $\hat{P}(e_{t-1})$ as $n(e_{t-1})$, a normalizing function which insures the posterior distribution $\hat{P}(e_t | e_{t-1})$ sums to 1.0 for all possible values of $e_t$.

Thus, assuming we can easily compute $\hat{P}(e_t, e_{t-1})$, we can then determine our current estimate of the true distribution, i.e. the posterior for any stage in the tree as $\hat{P}(e_t | e_{t-1})$. These revised posteriors are shown schematically for a simple 3 stage process in Figure 6-2.

In practice, we can build up the estimated distribution $\hat{P}(e_t | e_{t-1}, e_{t-2}, \ldots, e_1)$ at each node of the tree as we progress from period 1 on through the tree using Bayes theorem. We accomplish this in the traditional manner by assuming the posterior for any stage $t$ is simply given as the product of the previous best estimate for $t$ as our new prior estimate at the current stage, i.e. simply the posterior from the previous stage, times a subjective conditional distribution. Thus, the posterior distribution for stage $t$ is given by
Figure 6-2
The Conditional Event Tree for the Multistage Sequential Investment Problem
\[ P(e_t \mid e_{t-1} = \phi_k) = \frac{\hat{P}(e_{t-1} = \phi_k) \cdot g(e_t = \phi_k \mid e_{t-1})}{n(e_{t-1} = \phi_k)} \]

where the present prior \( \hat{P}(e_{t-1} = \phi_k) = \hat{P}(e_{t-1} = \phi_j) \), the posterior from the previous stage, and \( g(e_t = \phi_k \mid e_{t-1}) \) is a stage independent conditional distribution which encodes how much our prior estimate should be changed given the observed result at stage \( t-1 \).

In normal Bayesian terms,

\[ \hat{P}(\phi \mid z) = \frac{\hat{P}(\phi)g(z \mid \phi)}{n(z)} \]

The prior \( \hat{P}(e_{t-1} = \phi_k) \) is simply the posterior from the previous stage which has encoded all of the historical results to that point and was calculated in the same way. The function \( g(e_t = \phi_k \mid e_{t-1}) \) is the conditional distribution of event \( e_t \) occurring given the value of the previous stage \( e_{t-1} \). This is stage independent and represents how much confidence we place in an observed value — i.e., how much an observation will change your estimate of the true distribution. In most applications, this process is a well known and well behaved function.\(^1\) In our case, it must also be subjectively estimated.\(^2\)

---

\(^1\)The process generally suggests the function. For example, in traffic counting, a Poisson distribution is a natural distribution to use as the conditional.

\(^2\)This will perhaps provoke the greatest outburst from critics of the Bayesian approach. There are a number of ways of relaxing this assumption. One is to perform a sensitivity analysis over the impact of this conditional. A second, and we have explored this using some simple programs in a related study and described this formulation in Appendix D, is to incorporate a second level of uncertainty over the actual conditional we use. Thus we learn not only about the true distribution but about the conditional that characterizes the process as well.
6.3 Incorporating the Adaptive Mechanism into the Basic Multiperiod Algorithm

6.3.1 A Discrete Version: DECISN with Bayesian Revision

We have incorporated the Bayesian model of the previous section in the programs DECISN as a discrete process in keeping with the formulation of previous chapters. In line with the assumption of an earlier model [9] and Appendix D of this study, we let the conditional distribution \( g(z|\phi) \) be simply a function of \( (z-\phi) \):

\[
g(z|\phi) = h(z-\phi)\]

In essence, we are defining a conditional with a shifting axis depending on the true value of \( \phi \). \( \phi \) and \( z \) are measured on the same scales. If we assume the number of discrete values of both \( \phi \) and \( z \) is \( m \), then the range of \( (\phi-z) \) is from \((1-m)\) to \((m-1)\). For convenience, we add a scale factor and define the variable \( km \) where

\[
km = \phi-z + m \quad ; \quad km = 1, \ldots, k_{\text{max}}
\]

where \( k_{\text{max}} = (2 \cdot m) - 1 \)

Each probability distribution is defined over \((\phi-z)\) as a function of \( km \). For the 3 state variable distribution we have used for the majority of the experiments \((m=3)\), we require a conditional with

\[
k_{\text{max}} = (2 \cdot 3) - 1
\]

\[= 5\]

elements. We now turn to exploring the effects of the adaptive model for a variety of priors and conditionals.
6.3.2 Variations in the Subjective Stage Probabilities for Different Priors and Conditionals

The purpose of this section is to demonstrate the effect of repeated application of the Bayesian revision mechanism for the multistage problem. Using a uniform prior distribution of $p_k(\phi_k) = [.333 \ 0.333 \ 0.333]$ and a uniform conditional $g(e_t = $ $\phi_k | e_{t-1}) = [.2 \ 0.2 \ 0.2 \ 0.2 \ 0.2]$, the subjective probability distribution of the state variable at any stage remains the same; i.e. the probability of $\phi_k = \alpha_2$ at stage 3 is the same as the value at stage 1: in this case, it would be 0.333. In fact, by using a uniform conditional with any prior distribution, uniform or otherwise, there is no change. In other words, we are assuming that we learn nothing from the results of previous stages when we employ a uniform conditional. The procedure is therefore identical to the model in earlier chapters.

With the same prior but a fairly "strong" (peaked) conditional, $g(e_t = \phi_k | e_{t-1}) = [.05 \ 0.15 \ 0.6 \ 0.15 \ 0.05]$, the probability of $\phi_k = \alpha_j$ at stage $t$ is now dependent on where we are in the tree. For example, as shown in Figure 6-3, although the probability of $\phi_k = \alpha_1$ at stage $t=1$ is only 0.333, it has increased to 0.750 at stage 2 and 0.935 at stage 3 when the previous period observed results were $\phi_{k1} = \alpha_1$ and $\phi_{k2} = \alpha_1$.

Figure 6-4 demonstrates the effects of a non-uniform prior $[0.2 \ 0.6 \ 0.2]$ with both a strong and gentle conditional distribution.
Prior: .333, .333, .333
Conditional: .05, .15, .60, .15, .05

Figure 6-3
Variations in the Subjective Stage Probability Distributions
314
Stage: 1

3 2
.935 .058
.480 .480
.040 .040
.364 .364
.273 .273
.056 .056
.040 .040
.480 .480
.364 .364
.040 .040
.480 .480
.006 .006
.935 .935
.480 .480
.040 .040
.364 .364
.056 .056
.040 .040
.480 .480
.006 .006
.935 .935
Figure 6-4
Variations in the Subjective Stage Probability Distributions
Varying Conditionals

Stage: 1 2 3
The strong conditional produces the unbracketed numbers at each stage. The effects of a gentle conditional appear beneath these numbers in brackets. Note that for extreme branches of the tree \([\alpha_1, \alpha_1, \alpha_1]\) or \([\alpha_3, \alpha_3, \alpha_3]\) there are quite different effects occurring, depending on which conditional distribution is in use. Although we begin with a low probability estimate in stage 1 for both \(\alpha_1\) and \(\alpha_3\), by stage 3, if we are on a low or high spine, we have changed significantly the estimated probability distribution for the strong conditional. The other distribution affects the subjective posterior distribution in the same way but with much less impact. Thus, we can represent our degree of confidence in observations by a more or less peaked conditional learning distribution. For the uniform conditional case, probabilities do not change at different points in the tree and the computational technique of backwards recursion can be employed to reduce the computations (assuming no activity shift effects). With either adaptive learning or activity shift, we will need to employ the averaging-out and folding-back procedure of Chapter 3.0. With this discussion as background, we now turn to a presentation of the application of the (learning) adaptive version of the model to the investment problem of Chapter 5.0.
6.4 Computational Experience with the Learning Adaptive Model for the Project Design Capacity Expansion Example

6.4.1 Computational Results

In order to explore the effects of the learning adaptive model of this chapter on the transport investment problem, we first applied the model to the example of Section 5.3.5.1 — that is, without the activity shift model operative. Recall that in order to reduce computational expense, we restricted the number of alternative plans to 24, state variables to 3, and stages to 2 (5 and 10 years in length). Table 6-1 compares the results of Section 5.3.5.1 with those predicted using the same initial prior of \( P_k(\phi_k) = [0.333\ 0.333\ 0.333] \) but now using a conditional of \( g(e_t = \phi_k | e_{t-1}) = [0.05\ 0.15\ 0.6\ 0.15\ 0.05] \) at 3% population growth per year and a 5% discount rate.

Note that the full stochastic tree search using the learning adaptive model produces a solution (\( \Lambda_j^- \), no initial investment) different from both the deterministic solution (\( \Lambda_5^- \), build a 2 lane facility in location C) and the non-sequential static uncertainty solution (\( \Lambda_2^- \), expand the existing facility in a 2 level expansion).

Also recall that the non-sequential solution itself will be an improvement, in terms of expected net present value, over the optimal sequence for the mean state variable when evaluated in an expected value sense against all state variable levels.\(^1\) However, the strategy

\(^1\)We have only shown the net present value of this sequence, \( \Lambda_5^- \), for the mean state variable level. This should grossly overestimate the expected value of \( \Lambda_5^- \) when evaluated against all state variable levels using the expected value mechanism.
Table 6-1
Comparison of Sequential and Non-Sequential Solutions

<table>
<thead>
<tr>
<th>Type of Tree Search</th>
<th>Investment Sequence</th>
<th>Total Expected Net Present Value ($\times 10^6$)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deterministic-low demand</td>
<td>$\Lambda_1 \ \Lambda_2$</td>
<td>-1.598</td>
<td>$\alpha_1$ spine</td>
</tr>
<tr>
<td>Deterministic-mean demand</td>
<td>$\Lambda_5 \ \Lambda_5$</td>
<td>2.721</td>
<td>$\alpha_2$ spine</td>
</tr>
<tr>
<td>Deterministic-high demand</td>
<td>$\Lambda_5 \ \Lambda_7$</td>
<td>9.998</td>
<td>$\alpha_3$ spine</td>
</tr>
<tr>
<td>Non-Sequential Static Uncertainty</td>
<td>$\Lambda_2 \ \Lambda_4$</td>
<td>2.209</td>
<td>Full tree search but with no revision of probabilities</td>
</tr>
<tr>
<td>Sequential Dynamic Uncertainty</td>
<td>$\Lambda_1 \ \ast$</td>
<td>2.887</td>
<td>Full tree search with revision of probabilities</td>
</tr>
</tbody>
</table>

* - denotes conditional decision which depends on state variable observed at first stage.

318
<table>
<thead>
<tr>
<th>Type of Tree Search</th>
<th>Investment Sequence</th>
<th>Total Expected Net Present Value ($ \times 10^6$)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deterministic-low demand</td>
<td>$\Lambda_1 \ \Lambda_2$</td>
<td>$-1.598$</td>
<td>$\alpha_1$ spine</td>
</tr>
<tr>
<td>Deterministic-mean demand</td>
<td>$\Lambda_5 \ \Lambda_5$</td>
<td>$2.721$</td>
<td>$\alpha_2$ spine</td>
</tr>
<tr>
<td>Deterministic-high demand</td>
<td>$\Lambda_5 \ \Lambda_7$</td>
<td>$9.998$</td>
<td>$\alpha_3$ spine</td>
</tr>
<tr>
<td>Non-Sequential Static Uncertainty</td>
<td>$\Lambda_2 \ \Lambda_4$</td>
<td>$2.209$</td>
<td>Full tree search but with no revision of probabilities</td>
</tr>
<tr>
<td>Sequential Dynamic Uncertainty</td>
<td>$\Lambda_1 \ *$</td>
<td>$2.887$</td>
<td>Full tree search with revision of probabilities</td>
</tr>
</tbody>
</table>

* - denotes conditional decision which depends on state variable observed at first stage.
suggested by the learning adaptive version of the model is an even better solution since it recognizes the value of observing the first stage results before making a second stage decision. The improvement in expected net present value between the sequential vs. non-sequential solutions is approximately 30%, even though the sequential solution suggests no initial investment.

Note that it has also selected a more "flexible" initial investment alternative in terms of plan evolution (alternative \( \Lambda_1 \)) than either of the other two solutions. Alternative \( \Lambda_1 \) can easily reach all of the existing plans. If demand at the end of stage 1 turns out to be low, we can continue with the existing alternative or a slight expansion. If we had committed ourselves to a large scale decision in the first stage and demand turned out to be low, we would be faced with large scale excess capacity losses.

On the other hand, if demand turns out in fact to be high at the end of stage 1, we can still expand to a larger scale facility. The predicted optimal strategy, \( S^* = \{ \Lambda_1, R \} \), is shown in Figure 6-5. If demand is low at the end of stage 1, \( \alpha_1 \), the second stage investment is to expand the existing 4 lane facility to an improved alternative \( \Lambda_3 \). For a first stage mean level outcome, \( \alpha_2 \), the optimal decision is to build the new facility, the 2 lane in location G (alternative \( \Lambda_5 \)). And finally, if the \( \alpha_3 \) level occurs, the suggested decision is to build the new full 4-lane facility, \( \Lambda_9 \).
Optimal Strategy \( S^* = \left[ \begin{array}{c} \Lambda_1 : \Lambda_3 | \phi_{k,1} = \alpha_1 \\ \Lambda_5 | \phi_{k,1} = \alpha_2 \\ \Lambda_9 | \phi_{k,1} = \alpha_3 \end{array} \right] \)

Figure 6-5
The Optimal Investment Strategy for the Two-Stage Model
320
6.4.2 The Effect of the Learning Adaptive Model on Technology Adaptation

We now return briefly to the example of Section 5.3.3 which explored the effect of building adaptability directly into the alternative by buying right-of-way for expansion ahead of time. Recall that the investment alternatives were constrained to the following:

\[ \Lambda_1: \text{the existing alternative} \]
\[ \Lambda_5: \text{a new 2 lane with r.o.w. for 2 lanes} \]
\[ \Lambda_7: \text{expansion of } \Lambda_5 \text{ to 4 lanes} \]
\[ \Lambda_6: \text{a new 2 lane with r.o.w. for 4 lanes} \]
\[ \Lambda_8: \text{expansion of } \Lambda_6 \text{ to 4 lanes} \]

Therefore one alternative expansion set provided for no additional right-of-way until expansion occurred (\(\Lambda_5-\Lambda_7\)); the other provided for expansion by buying right-of-way for four lanes at the time the first stage, 2 lane investment occurred (\(\Lambda_6-\Lambda_8\)). \(\Lambda_6\) is therefore defined as more adaptable than \(\Lambda_5\).

The results are repeated in Table 6-2 for convenience and summarized here. For the low state variable spine \(\alpha_1\), the investment sequence selected was \(\Lambda_1-\Lambda_5\). For both the mean and high demand level, spines, \(\alpha_2\) and \(\alpha_3\), the adaptable alternative sequence, \(\Lambda_6-\Lambda_8\), was selected. A non-sequential static uncertainty run produced \(\Lambda_6-\Lambda_8\) as the unconditional strategy with an expected net present value of \(\$0.479 \times 10^6\) at a 10% discount rate.
Table 6-2

Tradeoffs between Technology Adaptation and the Adaptive Nature of Strategies

<table>
<thead>
<tr>
<th>Model Type</th>
<th>State Variable Level (x 10^-6)</th>
<th>Investment Strategy</th>
<th>Expected Net Present Value (x 10^6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deterministic</td>
<td>( \alpha_1 = .035 )</td>
<td>( \Lambda_1 )</td>
<td>( \Lambda_5 ) ( \Lambda_6 ) ( \Lambda_8 )</td>
</tr>
<tr>
<td></td>
<td>( \alpha_2 = .050 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \alpha_3 = .065 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Sequential</td>
<td>Stochastic</td>
<td>( \Lambda_6 )</td>
<td>( \Lambda_8 ) ( \Lambda_5 )</td>
</tr>
<tr>
<td>Sequential</td>
<td>Stochastic</td>
<td>( \Lambda_5 ) ( * )</td>
<td></td>
</tr>
<tr>
<td>(Learning Adaptive)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* - denotes conditional decision strategy
Turning to the learning adaptive model of this section, i.e. running the model with a non-uniform conditional \( g(e_t = \phi_k | e_{t-1}) \), we observe a change in the optimal investment pattern to a conditional strategy, \( \Lambda_5 - * \), which produces a different initial investment from the non-sequential case. If demand for the first stage turns out to be low, the investment need not be expanded beyond the 2 lane alternative \( \Lambda_5 \). For first stage results corresponding to the mean and high state variable levels, investment is expanded to \( \Lambda_7 \) in the second stage in both cases. The value of learning in this instance is an increase in total benefits from \( $.479 \times 10^6 \) to \( $.496 \times 10^6 \) for a change from a non-sequential static uncertainty solution, \( \Lambda_6 - \Lambda_8 \), to a sequential dynamic uncertainty investment strategy, \( \Lambda_5 - * \).

Thus, we see that a tradeoff exists between the flexibility provided by the conditional decision process and the adaptability of technology itself. In this instance, it is more beneficial to forego expansion savings of the adaptable alternative, \( \Lambda_6 \), by switching to an initial investment, \( \Lambda_5 \), which, apparently, is much less expensive if demand turns out to be low (i.e. we continue with \( \Lambda_5 \)) and yet can still expand to a four lane, \( \Lambda_7 \), which is nearly as good as the adaptable alternative expansion, \( \Lambda_8 \).

Again, the value of the stochastic sequential model is that it directly evaluates this trade-off through the sequential strategy comparisons. The extension to network investments will be discussed in Chapter 7.0. First, we turn to the results of variations in the prior and conditional distributions.
6.4.3 Effects of Variations in the Conditional Learning Function and Prior Distribution

To test the sensitivity of the selected investment strategy to more or less confidence in observed results and to varying initial conditions, we conducted a series of experiments varying both the prior and conditional probability distributions. Table 6-3 contains the results of these experiments.

The vertical columns represent variations in the selected strategy for different initial prior distributions and a constant conditional. The prior distributions used were:

1) uniform, P-I
2) strongly peaked, P-II, and
3) skewed-right, P-III

Rows represent variations in the best strategy for a change in the conditional from a strongly peaked distribution (C-1) through a uniform distribution (C-3) holding the prior distribution fixed.

Note that the mean state variable ($\bar{x}$) optimal sequence for all conditionals and priors is identical, since it is evaluated as if there were only one true value of demand with probability equal to 1.0. We include it for each case for comparison purposes only.

For the conditional used in the last section, conditional C-1, as we change from a uniform prior (P-I) to a peaked prior (P-II), we observe a drop in expected net present value from $2.887 \times 10^6$ to $2.654 \times 10^6$. For the former prior, with equal probability of high and low demand levels occurring, a strategy approach is considerably
Table 6-3
Variations in Strategy Selection due to Changed Priors and Conditionals

| Conditional Distribution | C-1: .05 |  | C-2: .15 |  | C-3: .20 |  |
|--------------------------|----------|  |----------|  |----------|  |
| Prior Distribution      |          |  |          |  |          |  |
| Investment Sequence     | ENPV     |  | Investment Sequence | ENPV     |  | Investment Sequence | ENPV     |
| x10^6                    |          |  | x10^6     |  | x10^6     |  |
| P-I .333 .333 .333      | $\bar{\alpha}$: $\Lambda_5$-$\Lambda_5$ 2.721 |  | $\bar{\alpha}$: $\Lambda_5$-$\Lambda_5$ 2.271 |  | $\bar{\alpha}$: $\Lambda_5$-$\Lambda_5$ 2.271 |
|  .20 .60 .20            | $\tilde{\alpha}$: $\Lambda_1$-$*$ 2.887 |  | $\tilde{\alpha}$: $\Lambda_2$-$\Lambda_4$ 2.210 |  | $\tilde{\alpha}$: $\Lambda_2$-$\Lambda_4$ 2.209 |
| P-II                    |          |  |          |  |          |  |
|  .10 .50 .40            | $\bar{\alpha}$: $\Lambda_5$-$\Lambda_5$ 2.721 |  | $\bar{\alpha}$: $\Lambda_5$-$\Lambda_5$ 2.721 |  | $\bar{\alpha}$: $\Lambda_5$-$\Lambda_5$ 2.721 |
|                          | $\tilde{\alpha}$: $\Lambda_5$-$*$ 2.654 |  | $\tilde{\alpha}$: $\Lambda_5$-$*$ 2.654 |  | $\tilde{\alpha}$: $\Lambda_5$-$*$ 2.654 |
| P-III                   |          |  |          |  |          |  |
|  .10 .50 .40            | $\bar{\alpha}$: $\Lambda_5$-$\Lambda_5$ 2.721 |  | $\bar{\alpha}$: $\Lambda_5$-$\Lambda_5$ 2.721 |  | $\bar{\alpha}$: $\Lambda_5$-$\Lambda_5$ 2.721 |
|                          | $\tilde{\alpha}$: $\Lambda_5$-$*$ 4.654 |  | $\tilde{\alpha}$: $\Lambda_5$-$*$ 4.654 |  | $\tilde{\alpha}$: $\Lambda_5$-$*$ 4.087 |

Notes:

--  - not computed
*  - conditional decisions
ENPV - expected net present value

325
more desirable since there is a higher probability for either high or low demands occurring than with P-II. In the latter case, an investment strategy, as opposed to a sequence, is selected but there is less expected loss associated at the extreme points because of the lower probabilities of extreme demand levels occurring.

Changing from the peaked prior to a skewed-right distribution, P-III, we observe a change in the optimal strategy to one which has alternative \( \Lambda_5 \) as the new investment with a much increased expected value. This is due to the larger weight placed on the consequences of investments at the high demand levels.

We now turn to examine the effects of variations in the conditional distribution. Lowering our degree of confidence in observed results by flattening out the conditional from C-1 through C-2 to C-3, we note a change from a conditional strategy selection for the first conditional to an unconditional sequence for the latter two for all priors. Conditional C-2 is a relatively flat distribution giving little difference in expected net present value (\( \$2.210 \times 10^6 \) to \( \$2.209 \times 10^6 \)) from the pure uniform (C-3). For the uniform and peaked prior, the strategy shifts from a conditional decision for C-1 to an unconditional strategy for C-3 that we discussed in the previous section. The initial investment in this case is different for the two conditionals.

For the skewed-right prior, however, the initial investment is identical for both conditionals C-1 and C-3, although there is
still a difference in expected value due to the desirability of conditional decisions when we employ a strong learning conditional.

6.5 Conclusions and Extensions

This chapter has explored the effect of a learning adaptive extension to the general sequential decision model investigated in Chapter 5.0. From the preliminary experiments conducted in this and the previous chapter, we can draw some conclusions as to the effects of such a model on both investment patterns and in terms of the computational requirements. The purpose of this section is to summarize these conclusions and outline some of the more obvious additional experiments needed to fully explore the effects of such a model.

The current chapter has presented the results of a series of experiments using the capacity expansion model of Chapter 4.0 and 5.0 and an extension for adaptively learning about the true underlying stochastic distribution of the demand parameters. The experiments were executed without incorporating the long-run supply-demand activity-shift model, for a number of prior and conditional distributions. In a positive sense, the most interesting result of the experiments was a shift from the non-sequential solution of Chapter 5.0 to a conditional decision, strategic investment pattern. In some cases, the initial decision of this strategy was different from both the initial investment of the optimal sequence predicted for the mean demand level and the non-sequential static uncertainty initial decision.
We also observed the effects of the adaptive technology example of Section 5.3.3 when a learning adaptive model was introduced. We found a change from the adaptable technology to the unadaptable technology as a first stage investment with the second stage investment conditional on the results of the first stage. We conclude that there is a very real tradeoff between adaptability of technology and the "adaptability" of the conditional decision strategy approach, and that the stochastic sequential model conveniently handles this calculation.

Using a variety of conditionals and priors, we also explored the effects of variations in initial conditions on the investment strategy. As our confidence in observed results changes from extreme confidence to a point where we place absolutely no value on observed demand, the solution approaches the non-sequential static uncertainty solution of the previous section.

Changes in priors in terms of peaking and skewing was shown to have a corresponding change in the optimal strategy and the expected net present value.

From this third major series of experiments, we conclude that a sequential decision framework can produce quite different investment solutions than those predicted by a deterministic model, and, in some cases, by those predicted by a non-sequential static uncertainty model.

More importantly, the initial investment chosen is exactly opposite in principle from what would be expected using the conclusions
reached by a number of related studies [10,11].

In general, for uncertain demand, these studies predict an initial investment of greater capacity than the deterministic model selects.\textsuperscript{1} Our model suggests that we build less capacity (actually, it suggests \textbf{no} capacity expansion for most of the cases) until we are sure about demand. The apparent discrepancy between these two results is not really discouraging - we are really talking about two different model formulations. On the one hand, these studies have used examples which assume a stationary and known probability distribution. Their decisions are therefore unconditional decisions which account for the relative opportunity losses of being under or over-capacity. Their results correspond to the non-sequential static uncertainty results we have reported. And since they deal with problems with large scale economies, their results simply imply that it is more costly to be under-capacity than over-capacity. Our results, on the other hand, recognize that there may be uncertainty over both demand and the parameters of that distribution. Thus, by delaying our decision to invest in the facility size suggested by either the mean demand level or by the non-sequential static uncertainty result, we retain the option of no investment if demand

\textsuperscript{1}It is interesting to note that their conclusions point to using a \textbf{smaller} discount rate in a deterministic model to obtain the same effect as the stochastic version. This is exactly opposite to the commonly suggested procedure of accounting for future uncertainty by placing less value on the future through an increased discount rate.
does in fact turn out to be low, and we can still invest in the higher capacity facility if demand turns out to be high.

This difference in results can be attributed in part to the particular shape of the cost functions, the losses associated with over- and under-capacity decisions, the degree of scale economies, and so on. The principal and overriding reason for this difference, however, is due to the use of historical information for updating demand estimates and employing a sequential decision framework on the one hand, and on the other, of ignoring this information which results in a non-sequential solution.

Before we can make any further definitive statements about either these results or the approach in general, however, there must be some further experimentation. The most obvious first set of experiments, aside from exploring the sensitivity of the procedure to different additional priors and conditionals, is to test the pruning rules and terminal functions of the previous chapter on the revised-distribution event tree. Because these probabilities will change for the same state variable sequence, with and without a conditional, the minimum pruning probabilities for each stage will eliminate quite different parts of the tree. However, in principle, it should still eliminate the least desirable parts of the tree, i.e. those sub-trces which have a small probability of occurring. Therefore the effect of pruning should be no different for either version of the model, although in general, we would expect different investments will be selected for these different versions.
A second obvious set of experiments is needed to test the adaptive adaptive [sic] version of the model. That is, we would like to test the model variations when there is both a tracking adaptive (the activity shift model is operational) and a learning adaptive mechanism. This additional set of experiments will produce both conditional decisions because of long-run supply-demand dependence, and conditional decisions because we are learning about the true state variable distribution. However, the effects of both these conditions operating simultaneously is difficult to predict, a priori.

Additionally, both Chapters 5.0 and 6.0 have ignored that the demand parameters may be in some sense correlated from one period to another. Although the result may be similar in effect to the learning adaptive mechanism of this chapter, it is quite different in principle. Accommodating such effects will involve extensions into the area of autocorrelated time-series models [12].

Finally, in addition to these simple experimental extensions to this work, we recognize the limitations of our simple models and recommend further research into making the procedure operational in a real world context. This will take the form of developing more realistic models, extending the approach to the network investment problem, and incorporating other than economic efficiency issues.
Chapter 7

SUMMARY AND CONCLUSIONS
Chapter 7
SUMMARY AND CONCLUSIONS

The research reported in this thesis has been concerned with the problem of making investment decisions for transport facilities when faced with a highly changing and uncertain future environment. The primary emphasis of the study has been to consider the transport investment decision as a time-related sequential investment problem recognizing the fact that investments are usually implemented as a series of staged sequential increments to an existing and fairly extensive system. Using the general stochastic sequential decision model as the basic framework for this problem, the study has focused on both approximation procedures which make the approach feasible, and the more general philosophical question of searching for flexible transportation systems which can adapt to a wide range of uncertain future demand conditions. The purpose of the present chapter is to present a summary of this research effort, the conclusions reached and some suggested future extensions for this work.

7.1 Summary of the Research Effort

One of the primary problems that transport investment is concerned with is the question of when, where, and how much capacity to add to existing and fairly extensive systems. Most of the research accomplished for or within the more recent transportation planning studies designed to answer these questions has focused on topics such as the following:

(1) predicting the demand for new technology

333
(2) modelling the interdependence between transport modes and other economic sectors

(3) analyzing the interactions of fairly diverse transport systems for some future design year.

Notably, very little effort to date has been devoted to what is probably the most significant aspect of planning large-scale durable investment facilities, i.e., recognizing both the fact that transport investments are implemented as staged, incremental additions to an existing system, and more importantly, that they are subject to very substantial future uncertain outcomes.

The purpose of this thesis has been to explore this problem of making investment decisions for transport facilities when faced with a highly changing and uncertain future environment. The unique nature of this study has been to recognize, in addition to the uncertain nature of future demand, the following characteristics of the transportation problem:

(1) links in a network are dependent on the connections with other links—a planning model must account for this topological connectedness when considering capacity additions in order to fully account for second-order "network effects" as well as the primary costs and benefits due to a proposed change.

(2) the demand for transportation is a derived demand that exists for achieving other than transport objectives—it is a function of both the socio-economic characteristics of the zones and of the characteristics of the
technology; planning models must be capable of handling a highly purpose, mode, and time of day differentiated set of demand functions and of capturing the dependency that exists between the supply of, and the demand for transportation capacity improvements.

(3) links on a network are highly dependent on other links simply because of financial constraints—because a dollar spent on one project cannot be spent on another, there can be simple financial dependencies between projects (known more generally as budget constraints).

(4) investments in transport facilities involve substantial future impacts and must be related to potential future investments—a planning model must be intertemporal in order to capture the multistage aspects of transport investments.

(5) and finally, there can also be severe timing interdependencies as well because of the interrelatedness of the socio-economic variables and transport investment capacity and accessibility—a planning model must be capable of accounting for dependencies of a longer-run nature between supply and demand.

The general equilibrium framework developed in a larger, related study effort is designed to handle these characteristics and provided the basic model structure for this study. A very general review of the literature showed that most existing planning studies can
be described in terms of this general equilibrium framework but only as special cases, i.e., they do not recognize all of the characteristics by ignoring some dependencies and/or approximating others.

In a more specific review of the literature, we found that all of the current approaches to time-staging of transport investment fall naturally into the two major categories of 1) project design and 2) program selection. The distinction between these two approaches is whether we are dealing with multiple projects or not. Program selection can be further subdivided into three categories which become progressively more complex by recognizing more and more of the above dependencies. These categories are:

1) Capital budgeting problems
2) Network flow problems
3) Activity shift problems

Within each of these categories, most of the emphasis has been on applying prescriptive optimization algorithms to fairly simplified problem structures or on massive complex simulation models which have no concern for the search problem of a combinatorially large solution space. The one exception has been the recent application of discrete optimization techniques of branch and bound and dynamic programming to transport investment network flow problems. Even these approaches, however, have required some extremely strong assumptions about the problem framework.

The specific contribution of our work has been to extend a model based on the general equilibrium framework to include the efforts
of future demand uncertainty and the sequential nature of implementation. Because of the complexity of general equilibrium models—which include all of the above constraints—and because of the changing nature of specific demand and supply models, most of the effort of our research has been of an exploratory nature. That is, we have not been concerned with specific investment policy decisions for a particular problem environment or the appropriateness of a particular demand model. Rather, we have been more concerned with the philosophy of planning in the face of uncertainty and the specific application of the sequential decision model to the general equilibrium framework of transport planning. The approach we have taken is general enough that the nature of the specific demand or supply models or equilibrating procedure is not central to the framework.

More specifically, after exploring both the equilibrium framework and the general sequential decision model in Chapter 2.0, we developed a discrete, stochastic capacity expansion model for analyzing transport investments as time-staged conditional strategies. In developing this specific model, our purpose was threefold: first, we wanted to be able to explore the issues of time-staging in a fairly simple setting and yet still be able to explore the implications of time and uncertainty on investment decisions for a realistic problem. Thus we chose a single mode investment problem (highways) but used the demand parameters of a recently developed static demand model as the uncertain variables. Using this model, we can compare differences in investment patterns for deterministic vs. stochastic conditions, for dynamic vs. static uncertainty.
conditions, for economies of scale vs. staging options, and so on, for a wide variety of conditions with very few restrictions on the shape of demand functions, the nature of costs, etc.

Second, since we wanted to be able to have an environment which closely resembles that of many large transportation studies, we developed the model in terms of the general equilibrium framework including long-run supply demand dependencies as well. Using the general equilibrium framework as a sequential decision model, however, results in a decision tree in extensive form of extreme dimensions. A straightforward application of the averaging-out and folding-back global search rule will be hardpressed to produce solutions for most problems. Our second major objective, therefore, was to develop approximating procedures—pruning rules and terminal evaluation functions—which help in making application of the sequential decision model in extensive form feasible for the general equilibrium transport planning framework.

Third, very rarely will we know exactly the probability distribution of the uncertain variables. The general sequential decision model should recognize that the information about the underlying probability distribution can also be to some degree uncertain. To handle this problem, we extended our transport planning model to include an adaptive mechanism (using a Bayesian updating procedure) for the sequential decision model. Incorporating this procedure in our basic model allows us to model investment decisions as they will be made in real life—in a sequential adaptive manner.
To accomplish these objectives, we implemented the time-staging model as a series of computer programs (DECISION) and conducted some initial experiments to explore the concept and the feasibility of the sequential decision framework as applied to transport investment planning. The programs have been developed as three major sets of routines. The first is the general decision tree routine for exploring and evaluating the utility of a variety of staged investment strategies under a stochastic demand structure. These routines include the standard averaging-out and folding-back expected value procedure as well as a number of specific pruning rule and terminal evaluation functions for pruning parts of the tree. The second major set of programs is the simulation model itself (MODEL) for predicting the equilibrium between supply and demand at a stage, and for predicting socio-economic growth changes (demand shifts), for the simple case of project design or single link analysis. It is with these two sets of routines that we have conducted our initial experiments. And finally, we developed a third series of routines (NETWK) for simulating the consequences of large-scale network equilibrium at a stage. These routines, at the present time have not been fully integrated with the DECISION model as a time-staged investment planning model.

7.2 Conclusions

The following three points summarize the general conclusions reached from the research reported on in this study.

1. First of all, from our survey of the literature we can conclude that there is no single computational formulation that is
appropriate for the transport investment planning problem. None of the existing computational techniques such as the presently popular mathematical programming techniques or even simulation techniques in their present form, are computationally feasible for the general transport investment problem under uncertainty.

2. We have hypothesized that the general sequential decision framework is necessary and a desirable addition to the transport investment problem because of the uncertainty over future outcomes and, more importantly, because of the sequential nature of transport investment implementation.

3. The uniqueness of the transport investment problem makes even the statistical decision theory framework, or even the more general sequential decision framework far from an operational procedure. Therefore, in order to apply this framework, it is necessary to develop a fairly elaborate set of heuristics.

We have recognized this problem and developed a computational framework which will serve as an experimental medium for exploring properties of these heuristics and for exploring basic issues of staging investments under uncertainty by using the heuristics on a number of example problems.

On the basis of limited exploration of these example problems, we can conclude the following as to the usefulness of the sequential decision model approach for investment planning, and on the value of the heuristic approximating procedures:
1. First, planning in terms of a sequential decision or strategic approach can lead to quite significant differences in the initial investment pattern and differences in expected value from a deterministic approach. Additionally, we conclude that there is a difference between recognizing uncertainty and both recognizing uncertainty in a sequential problem, i.e., there are significant differences between non-sequential and sequential decision problems in the face of uncertainty.

2. A learning adaptive model will also change the investment pattern chosen from one selected using a non-sequential static uncertainty framework which assumes the probability distribution is known exactly. The extent of this change will depend on the degree of confidence expressed in the underlying stochastic environment represented by the shape of the conditional learning distribution.

3. There are a number of useful pruning rules and terminal evaluation functions which make application of the general sequential decision model to transport planning feasible. The experiments conducted indicated that the use of myopic selection procedures coupled with stage aggregation techniques and selected probability spines can be applied in a relatively straightforward manner. In small network problems, the use of these techniques
can approximate large portions of the decision tree fairly accurately. In large network problems, the use of these techniques should be even more appropriate.

Stated in summary, the primary conclusion reached from exploring the application of the sequential decision framework is that transportation planning should be carried out in such a framework based on both a priori considerations and on experimental results, and that it is potentially feasible to do so.

7.3 Extensions and Directions for Future Research

Based on our survey of the literature and on experience with the model, there are a great many directions in which to profitably extend this research. These extensions can be divided into two distinct but related areas: (1) those relating to extending and improving the computational techniques of the specific model employed in this thesis; the purpose being to test and improve on the efficiency of the algorithm and its variants for a wide variety of problem types, and (2) those extensions relating to the larger problem of planning transportation improvements; including new as well as existing technology, in the strategic sense used in this thesis, and questions of equity and social implications of transportation investments.

First, the specific extensions to the current model:

(1) The experiments reported on in Chapters 5.0 and 6.0 are clearly only preliminary in terms of fully evaluating the various pruning rules and terminal functions. Besides extending the experiments to more fully test their characteristics for a number of project design
problem types, there are a great number of variants of the existing rules, singly and in combination with other rules, that should be tested and evaluated in a variety of problem contexts.

(2) A second obvious extension would be to fully integrate the network routines with the decision tree programs to explore the properties of the pruning rules and terminal functions in this environment. A priori, we would expect that as the number of potential investment alternatives increased (due to the combinatorial effects of multiple programs), the benefits of a flexible search procedure using heuristic pruning rules and terminal evaluation functions would also increase dramatically. However, although the current single link programs recognized the network effects of alternative paths, we anticipate that there will be some properties of the pruning rules and terminal evaluation functions unique to the larger network problem. For example, the use of an horizon approximation procedure similar to the approaches of Roberts [1], Consad [2], Weingartner [3], and others would tend to cut down dramatically on the number of alternative sequences to evaluate and be much more useful for the network case.

Incorporating the constraints of multi-period budgets for the network problem is yet another area that could usefully be explored. The constraints of budgets should also have quite different effects on the selection procedures of the model.

(3) A third extension would be to incorporate the already developed, partial optimization procedures of Loubal [4], Ridley [5], Chan [6], etc., discussed in Chapter 2.0 for selecting good alternatives at
any stage. Additionally, incorporating branch and bound and other tree search techniques [7] or dynamic programming as partial optimization schemes at any stage, or for a sub-tree, are other possible approaches for improving stage selection procedures that could usefully be explored as improvements to the current search procedures. Additionally, the relationship of other direct search procedures should be investigated for their applicability and usefulness [8].

(4) On a more experimental basis, there are a number of techniques not yet fully developed that could also be extremely useful in heuristically reducing the search space. To date, they have been directed mainly to single period problems. They too should be explored for their applicability to the time-staging problem. For example, network aggregation [9]—the technique of using a collapsed network in order to estimate improvement benefits—should significantly reduce the simulation time at each stage—the primary factor which will seriously hamper any transportation time-staging model. In a similar manner, the use of other models, such as linear programming flow models in the spirit of the hierarchically structured multi-level model approach of Manheim [10], Bruck, Manheim, and Shuldiner [11], is another possible approach. Although other areas such as water resources (Males [12]) have attempted to evaluate the usefulness of mathematical programming models in directing the search for larger scale simulation models, and some preliminary work has been carried out in a related transportation research study effort [13], there seems to be no specific work in the transportation literature that can make other than very general comments as to the usefulness of this approach.
(5) And fifth, a final extension would be to expand the problem environment. For example, the simulation models themselves should be expanded to include a number of different modes in order to explore the effects of uncertainty on transport systems with extremely different operating characteristics and costs. An appropriate extension would be to build the decision model characteristics into an existing, realistic planning model such as the current model system developed at M.I.T. [14]. We would also like to extend the analysis to include more than one uncertainty factor, allowing demands, growth factors, and supply uncertainties to be handled by the model. It is in this kind of environment that the flexibility of the search procedures, the ability to select high and low probability spines, to stage aggregate and so on will show their real usefulness.

In terms of the second type of extension, i.e., the larger problem of planning transportation improvements recognizing that questions of economic efficiency are only a part of the process, there is considerable work to be done in adjusting the sequential decision framework to accommodate techniques which are more appropriate in recognizing the social and political implications of investment policies; for example, techniques such as the standard cost-effectiveness approaches [15], the more recent goal fabric approach [16], and so on.

Additionally, our literature survey showed a variety of time-staging models ranging from project design through capital budgeting and growth models which probably all have their appropriate place in planning investments. A very real need exists for relating these areas--
for example, when can we design a project in absence of network effects, how can capital budgeting help to choose project investments even though it ignores true timing interdependencies and assumes costs and benefits are exogenously specified, what is the role of capital budgeting in relationship to network growth models, etc. The sequential decision framework could be usefully extended to each of these areas, and integrated with other planning procedures.

And finally, there remains the whole question of new technology. It was an initial interest in new technology and the uncertainty that surrounds such decisions that was the original motivation for exploring the problem of investment time-staging and the application of the sequential decision framework. However, it has only been in the last few years, since the beginning of the large-scale urban transportation studies, the urban freeway crisis and the recognition of regional transportation shortcomings, that any serious effort has been devoted to basic research on new technology. Present forms of new technology—the dual mode technology in its many shapes and forms, the technology of the dial-a-ride system, and so on—will soon be in the full-scale experimental test stage. It is in this area that we need to carefully explore these new modes in terms of the sequential decision framework; which modes can best adapt to changing and uncertain conditions of the future and which will lock us into dead-end technologies or be extremely costly to change. It is here that the building of adaptable and flexible technologies that can evolve to match changing conditions would appear to be extremely desirable.

In summary, it will not be a better computer algorithm or a
better set of heuristics that will solve the problems of transportation. However, the philosophy of such models and their use can help us to make better decisions concerning what investments, where and when to make them. It is in this spirit that the research of this study was undertaken. We seriously hope investment planning of new technology will incorporate the sequential decision philosophy in principle, if not in actual practice.
REFERENCES

Chapter 1.0


[18] Hanson, loc. cit.


[38] Northeast Corridor Transportation Project, _loc. cit._


[41] Manheim, _et al., op. cit._, p. 252.

[42] Hanson, _loc. cit._


[50] Raiffa, loc. cit.


Chapter 2.0


[20] Ruiter, op. cit., pp. 204-244.
[21] Wohl and Martin, op. cit., p. 245


[57] Muhich, loc. cit.

[58] Russel, loc. cit.


[60] Tarplay and Drake, loc. cit.

[61] Thygeson, loc. cit.


357


Chapter 3.0


[22] Bergendahl, G., loc. cit.


[38] Raiffa, Howard, and Robert Schlaifer, APPLIED STATISTICAL DECISION THEORY, Boston, Mass.: Division of Research, Graduate School of Business, Harvard University (1961).


364
Chapter 4.0


365


Chapter 5.0


Chapter 6.0


Chapter 7.0


369


BIOGRAPHICAL NOTE

Wayne M. Pecknold was born on August 3, 1940 in Victoria, British Columbia. He received the Bachelor of Science degree in Civil Engineering at Michigan State University in June 1963, and Master of Science from the Massachusetts Institute of Technology in Civil Engineering in June 1965.

During his doctoral program at M.I.T., the author was an instructor in the department assisting in the transportation system analysis courses, an introductory computer course, and a course in decision theories for engineers and planners. Currently, he assists in the introductory core transport systems analysis course and teaches the second course in this sequence.

His research at M.I.T. has been in the general area of planning and analysis of transport systems with special emphasis on computer applications of long-run investment planning problems and new technology for urban transportation. Currently, in addition to his doctoral program, the author has been concerned with the development of the analysis and prediction of costs for PROJECT CARS, or Dial-a-ride system. His doctoral program at M.I.T. has focussed on three areas: transport systems planning, operations research and systems analysis, and economics. The author is a member of The Institute of Management Sciences, the Transportation Research Forum, the Operations Research Society of America, and is also a member of the Committee on Future Concepts of the Highway Research Board.

In September of 1961, he married the former E. Jean Arnott of Vancouver, B.C. They have three children; Kristen, Rand and Brett.

PROFESSIONAL EXPERIENCE

Consulting

Consultant, City of Cambridge – Preparation of Model Cities Grant application, Transportation Section, Spring 1967.


Teaching and Research

1966 to Present – Instructor, Transportation Systems Division, Department of Civil Engineering, M.I.T.

Research Assistant, Transportation Systems Division, Department of Civil Engineering, M.I.T., 1963-66.


Bayesian Decision Theory Learning Model for Highway Location, Universidad de Los Andes, Bogota, Colombia, Inter-American Program, M.I.T., 1964.


AWARDS AND HONORARY SOCIETIES:

Recipient – Tau Beta Pi Award for Outstanding Freshman Engineer, Michigan State University, 1959.

Tau Beta Pi
Chi Epsilon
Society of Sigma Xi
Phi Eta Sigma

PUBLICATIONS


PUBLICATIONS (Cont'd.)

Contributing author, Hanson, Mark E. (Editor), PROJECT METTRAN, Cambridge, Massachusetts: M.I.T. Press (November 1966).


APPENDIX A

NOTATION AND SYMBOLS

A specification of activity system; population, economic trends, etc.

A_i accessibility of zone i

A65,A75 specification of the activity system in 1965; in 1975

ACCPV(I,KK) the accumulated net present value for discount rate I up to and including stage KK

a(v) loss function of not adapting at volume level v

a_m action m

\bar{B}(\sigma_i) the total discounted expected net present value of initial investment \sigma_i

B(\sigma_i) the total discounted expected net present value of initial investment \sigma_i

C_k total cost of alternative k

c_{ij},P_{ij} total perceived price or cost of travel between zones i and j

c_t(\Lambda_{kt}) marginal cost of expansion to \Lambda_{kt} in period t

D,D(L,A) generalized demand function

d_t discount factor applied to year t using interest rate \rho

E_j,E_i employment located in zones i and j

E_{-t} sequence of events occurring up to stage t

e_{-t} event occurring at time t represented by a specific value of the state variable vector \phi_{kt}

ENPV expected net present value

1Symbols are presented in alphabetical order; first those of standard symbols, then those of Greek alphabet.
\( F(\Lambda_{mt}, \phi_{kt}) \) flow pattern resulting at time \( t \) from plan \( \Lambda_{mt} \) and state variable \( \phi_{kt} \).

\( F(V,L) \) general flow pattern of a specific volume and level-of-service

\( f_{ij} \) frequency of service between zones \( i \) and \( j \)

\( f \) accessibility factor

\( G(\Lambda_{mt}) \) gross benefits at stage \( t \) resulting from plan \( \Lambda_{mt} \)

\( \tilde{G}(\Lambda_{mt}, \phi_{kt}) \) gross benefits at stage \( t \) resulting from plan \( \Lambda_{mt} \) and state variable \( \phi_{kt} \)

\( g(z|\phi), h(z-\phi) \) conditional probability distribution of observing result \( z \) given true state variable level is \( \phi \)

\( H_{t-1} \) history of observed results to stage \( t-1 \)

\( h(z-\phi) \) conditional probability density function with a shifting mean value

\( I_k \) total facility investment cost

\( K(\xi_{mt}) \) total capital cost of investment program \( \xi_{mt} \) in period \( t \)

\( KK \) time period KK

\( k \) alternative state variable level

\( k_{\text{max}} \) maximum number of conditional distribution elements

\( L \) level of service vector

\( L_m \) economic life of plan \( m \)

\( L_i(t), L_j(t) \) decision variable for link between zones \( i \) and \( j \) (or link \( j \)) in period \( t \)

\( M(\Lambda_{mt}) \) total annual maintenance cost of plan \( \Lambda_{mt} \)

\( m \) alternative plan index \( m=1, \ldots, NALT \)

\( N, \hat{N} \) horizon period

\( n \) maximum number of link investments

\( n(z) \) normalizing function

\( \text{NPV} \) net present value
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$o_{ij}$</td>
<td>total operating cost of travel between zones $i$ and $j$</td>
</tr>
<tr>
<td>$P_{i,j}$</td>
<td>population of zones $i,j$</td>
</tr>
<tr>
<td>$\hat{\phi}(\phi</td>
<td>z)$</td>
</tr>
<tr>
<td>$\hat{\phi}(\phi)$</td>
<td>subjective prior distribution</td>
</tr>
<tr>
<td>PMIN(t)</td>
<td>minimum pruning probability cutoff</td>
</tr>
<tr>
<td>POST(KK)</td>
<td>posterior distribution at stage $KK$</td>
</tr>
<tr>
<td>PROD(KK)</td>
<td>partial product estimate of future from stage $KK$ on</td>
</tr>
<tr>
<td>PVD(KK,I)</td>
<td>present value (discounted at stage $KK$ with interest rate $I$)</td>
</tr>
<tr>
<td>PVU(KK)</td>
<td>present value (undiscounted) at stage $KK$ with interest rate $I$</td>
</tr>
<tr>
<td>$p_{ij}c_{ij}$</td>
<td>total perceived price of travel between zones $i$ and $j$</td>
</tr>
<tr>
<td>$p_k(\phi_{kt})$</td>
<td>probability distribution of state variable $\phi_{kt}$</td>
</tr>
<tr>
<td>$p''(\phi</td>
<td>z,e)$</td>
</tr>
<tr>
<td>$p'(\phi)$</td>
<td>prior probability density function</td>
</tr>
<tr>
<td>$q_k$</td>
<td>volume of flow on link $k$</td>
</tr>
<tr>
<td>$R_k$</td>
<td>total right-of-way cost of link $k$</td>
</tr>
<tr>
<td>$S_t$</td>
<td>investment vector showing link implementation times</td>
</tr>
<tr>
<td>$S^*$</td>
<td>optimal strategy; series of conditional investment plans</td>
</tr>
<tr>
<td>STACU(KK)</td>
<td>stage accumulated net discounted present value for stage $KK$</td>
</tr>
<tr>
<td>$T_\Lambda$</td>
<td>state transformation matrix</td>
</tr>
<tr>
<td>$T$</td>
<td>set of available transportation options; technologies, networks, vehicles and operating policies</td>
</tr>
<tr>
<td>$t$</td>
<td>annual time index $t=1,...,N$</td>
</tr>
<tr>
<td>$t_j$</td>
<td>timing of investment alternative $j$</td>
</tr>
<tr>
<td>UBEST(KK+1)</td>
<td>total future expected terminal value of the current alternative at stage $KK$ assuming an optimal strategy from $KK$ on</td>
</tr>
</tbody>
</table>
\( v_{ij}^m \)  
interzonal demand function for mode \( m \) between zones \( i \) and \( j \)

\( v_{ij}, q_{ij} \)  
volume of flow between zones \( i \) and \( j \)

\( v_k, q_k \)  
volume of flow on link \( k \)

\( w_{ij} \)  
expected accident costs of travel between zones \( i \) and \( j \)

\( x_k \)  
average user tax for link \( k \)

\( x_{ij} \)  
vehicle tax for the trip between \( i \) and \( j \)

\( y_i \)  
average per capita income of zone \( i \)

\( z \)  
set of results possible from experiment \( e_k \)

\( \alpha_0, \alpha_1, \alpha_2 \ldots \alpha_m \)  
set of demand function parameters

\( \bar{\alpha} \)  
mean state variable level

\( \bar{\alpha}_k \)  
state variable spine \( k \); a history of state variable levels over time

\( \beta_{mn} \)  
demand function parameter

\( \beta \)  
elasticity of demand with respect to price variable

\( \Gamma_k(\Lambda_{mt}, \phi_{kt}) \)  
terminal evaluation function \( k \)

\( \gamma_{mt} \)  
equivalent annual cost of alternative \( m \)

\( \Delta a_{i} \)  
change in accessibility of zone \( i \) from period \( t \) to \( t+1 \)

\( \delta_{mn} \)  
demand function parameter

\( \zeta \)  
budget constraint

\( \eta_k(\Lambda_{mt}, \phi_{kt}) \)  
horizon terminal evaluation function

\( \theta_{mn} \)  
demand function parameter

\( \Lambda_{mt} \)  
alternative plan \( m \) at stage \( t \)

\( \nu \)  
value of time

\( \xi_{mt} \)  
program of investments \( m \) occurring in stage \( t \)

\( \rho_t \)  
interest rate for period \( t \)
\( \sigma_i \)  
first stage investment program

\( T \)  
total path probability vector to stage \( t \)

\( \phi_{kt} \)  
state variable level \( k \) occurring at time period \( t \)

\( \psi_j \)  
scale of investment \( j \)
## APPENDIX B

### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>The Relationship of Project Design to the Overall Network Structure</td>
<td>20</td>
</tr>
<tr>
<td>1-2</td>
<td>The Flexibility of Technology</td>
<td>29</td>
</tr>
<tr>
<td>1-3</td>
<td>The Flexibility of Initial Network Programs</td>
<td>31</td>
</tr>
<tr>
<td>1-4</td>
<td>The General Sequential Decision Model in Extensive Form</td>
<td>35</td>
</tr>
<tr>
<td>2-1</td>
<td>Simple Equilibrium</td>
<td>55</td>
</tr>
<tr>
<td>2-2</td>
<td>Changing Equilibrium Due to System Improvement</td>
<td>56</td>
</tr>
<tr>
<td>2-3</td>
<td>Extending the Supply-Demand Paradigm to Multiple Time Periods</td>
<td>58</td>
</tr>
<tr>
<td>2-4</td>
<td>Basic Prediction Models</td>
<td>60</td>
</tr>
<tr>
<td>2-5</td>
<td>Multiple Scales of Transport Investment</td>
<td>66</td>
</tr>
<tr>
<td>2-6</td>
<td>Two Typical Investment Sequences for a Single Demand History</td>
<td>69</td>
</tr>
<tr>
<td>2-7</td>
<td>Intra-Temporal Demand Functions</td>
<td>72</td>
</tr>
<tr>
<td>2-8</td>
<td>Technology Supply Functions</td>
<td>75</td>
</tr>
<tr>
<td>2-9</td>
<td>Capacity Expansion Cost Paths</td>
<td>78</td>
</tr>
<tr>
<td>2-10</td>
<td>Equilibrium with the Average Variable Cost Function</td>
<td>80</td>
</tr>
<tr>
<td>2-11</td>
<td>Willingness-to-pay as a Measure of Benefits</td>
<td>83</td>
</tr>
<tr>
<td>2-12</td>
<td>Forming the Aggregate Supply Function for Two Parallel Links</td>
<td>86</td>
</tr>
<tr>
<td>2-13</td>
<td>Uncertainty Represented by Multiple Demand Functions</td>
<td>90</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>2-14</td>
<td>Single and Multistage Decision Trees</td>
<td>92</td>
</tr>
<tr>
<td>2-15</td>
<td>A Taxonomy of Investment Problems</td>
<td>97</td>
</tr>
<tr>
<td>3-1</td>
<td>The Transport Investment Problem Represented in Extensive Form</td>
<td>120</td>
</tr>
<tr>
<td>3-2</td>
<td>Alternative Network Plan ( \Lambda_m ) Describing Current and Potential Link Investments</td>
<td>122</td>
</tr>
<tr>
<td>3-3</td>
<td>Determining Equilibrium Flows</td>
<td>129</td>
</tr>
<tr>
<td>3-4</td>
<td>The Effects of Demand Parameter Variations</td>
<td>131</td>
</tr>
<tr>
<td>3-5</td>
<td>Flow Chart of the General Tree-Tracing Algorithm</td>
<td>154</td>
</tr>
<tr>
<td>3-6</td>
<td>Interpolating between Stages and Accumulating Net Present Values</td>
<td>159</td>
</tr>
<tr>
<td>3-7</td>
<td>The Detailed Expected Value Calculations</td>
<td>160</td>
</tr>
<tr>
<td>3-8</td>
<td>An Example Decision Tree for Comparing Search Efficiencies</td>
<td>163</td>
</tr>
<tr>
<td>3-9</td>
<td>The Decision Tree for Comparing Decision Analysis and Exhaustive Enumeration</td>
<td>165</td>
</tr>
<tr>
<td>4-1</td>
<td>Approximating Parts of the Decision Tree by a Sub-Problem ( p_k' )</td>
<td>173</td>
</tr>
<tr>
<td>4-2</td>
<td>Pruning Rules and Related Terminal Evaluation Functions</td>
<td>178</td>
</tr>
<tr>
<td>4-3</td>
<td>Embedded Terminal Evaluation Functions</td>
<td>181</td>
</tr>
<tr>
<td>4-4</td>
<td>Comparison of Myopic and Global Rules</td>
<td>185</td>
</tr>
<tr>
<td>4-5</td>
<td>Discrete Approximations to the State Variable Distribution</td>
<td>187</td>
</tr>
<tr>
<td>4-6</td>
<td>Selected Mean State Variable Spines</td>
<td>189</td>
</tr>
<tr>
<td>4-7</td>
<td>The Horizon-Approximation Sub-Problem</td>
<td>191</td>
</tr>
<tr>
<td>4-8</td>
<td>Flow Chart of DECISN's Tree-Tracing Algorithm</td>
<td>193</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>5-1</td>
<td>A Project Design Example</td>
<td>211</td>
</tr>
<tr>
<td>5-2</td>
<td>The Link Supply Functions</td>
<td>213</td>
</tr>
<tr>
<td>5-3</td>
<td>Forming the Aggregate Supply Function for Two Parallel Links</td>
<td>216</td>
</tr>
<tr>
<td>5-4</td>
<td>The Range of Demand over Time for Uncertain State Variable $\alpha_0$</td>
<td>221</td>
</tr>
<tr>
<td>5-5</td>
<td>Predicting Shifts in Population Due to Changing Accessibility</td>
<td>224</td>
</tr>
<tr>
<td>5-6</td>
<td>The General Strategy for Activity System Predictions</td>
<td>225</td>
</tr>
<tr>
<td>5-7</td>
<td>The Measure of Gross Benefits: Willingness to Pay</td>
<td>228</td>
</tr>
<tr>
<td>5-8</td>
<td>The Adaptability of Technology</td>
<td>246</td>
</tr>
<tr>
<td>5-9</td>
<td>Optimal Investment Sequences for Alternative Demand Spines</td>
<td>255</td>
</tr>
<tr>
<td>5-10</td>
<td>Computation Time for Various Decision Tree Sizes</td>
<td>294</td>
</tr>
<tr>
<td>6-1</td>
<td>The Basic Bayesian Decision Tree</td>
<td>305</td>
</tr>
<tr>
<td>6-2</td>
<td>The Conditional Event Tree for the Multistage Sequential Investment Problem</td>
<td>310</td>
</tr>
<tr>
<td>6-3</td>
<td>Variations in the Subjective Stage Probability Distributions</td>
<td>314</td>
</tr>
<tr>
<td>6-4</td>
<td>Variations in the Subjective Stage Probability Distributions - Varying Conditionals</td>
<td>315</td>
</tr>
<tr>
<td>6-5</td>
<td>The Optimal Investment Strategy for the Two-Stage Model</td>
<td>320</td>
</tr>
<tr>
<td>D-1</td>
<td>Illustration of Use of $\beta$</td>
<td>D-4</td>
</tr>
<tr>
<td>D-2</td>
<td>Logic of the Posterior Computations</td>
<td>D-8</td>
</tr>
<tr>
<td>D-3</td>
<td>Logic of the Computations for the Two-Dimensional Posterior</td>
<td>D-12</td>
</tr>
</tbody>
</table>
APPENDIX C

LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>The Potential Benefits and Costs of a Transportation Investment</td>
<td>82</td>
</tr>
<tr>
<td>3-1</td>
<td>Potential Network Plans for a Set of n Links</td>
<td>123</td>
</tr>
<tr>
<td>3-2</td>
<td>The Link Investment Matrix</td>
<td>125</td>
</tr>
<tr>
<td>3-3</td>
<td>Compact Notation for the Link Investment Matrix</td>
<td>126</td>
</tr>
<tr>
<td>3-4</td>
<td>A Typical State Transformation Matrix, $T_A$</td>
<td>127</td>
</tr>
<tr>
<td>3-5</td>
<td>The Combinatorial Effects of Multi-Period Models</td>
<td>135</td>
</tr>
<tr>
<td>5-1</td>
<td>The Plan Matrix for the Example Problem Showing 44 Distinct Plans Formed by Various Combinations of Projects</td>
<td>217(a)</td>
</tr>
<tr>
<td>5-2</td>
<td>Population and Income Data</td>
<td>219</td>
</tr>
<tr>
<td>5-3</td>
<td>Demand Parameter Values</td>
<td>220</td>
</tr>
<tr>
<td>5-4</td>
<td>Technology Characteristics</td>
<td>230</td>
</tr>
<tr>
<td>5-5</td>
<td>Comparison of Absolute vs Comparative Advantage</td>
<td>236</td>
</tr>
<tr>
<td>5-6(a)</td>
<td>Investment and R.O.W. Costs for the Adaptable Technology Example</td>
<td>249</td>
</tr>
<tr>
<td>5-6(b)</td>
<td>Results</td>
<td>249</td>
</tr>
<tr>
<td>5-7</td>
<td>Changes in Investment Sequences for Variations in the Cost of Adaptation</td>
<td>250</td>
</tr>
<tr>
<td>5-8</td>
<td>Optimal Investment Sequence for Selected Demand Level Spines</td>
<td>253</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>5-9</td>
<td>Results of Optimal Sequences at Other Demand Levels</td>
<td>256</td>
</tr>
<tr>
<td>5-10</td>
<td>Comparison of Deterministic Non-Sequential and Sequential Solutions for the Two Stage Model</td>
<td>262</td>
</tr>
<tr>
<td>5-11</td>
<td>The Effect of Accessibility on the Investment Sequence</td>
<td>267</td>
</tr>
<tr>
<td>5-12</td>
<td>Testing the Sensitivity of Investment to Elasticity of Demand</td>
<td>271</td>
</tr>
<tr>
<td>5-13</td>
<td>Results Comparing the Myopic and Global Rules for 2 and 3 Stage Models</td>
<td>278</td>
</tr>
<tr>
<td>5-14</td>
<td>Results of the Stage Aggregation Function</td>
<td>281</td>
</tr>
<tr>
<td>5-15</td>
<td>Effects of Variations in Number of Probability Classes</td>
<td>283</td>
</tr>
<tr>
<td>5-16</td>
<td>Pruning Probability Variations</td>
<td>285</td>
</tr>
<tr>
<td>5-17</td>
<td>Global Solution for Comparison with the Myopic Search Rule and Stage Aggregation</td>
<td>286</td>
</tr>
<tr>
<td>5-18</td>
<td>Results with Pruning Rule Set I and the State Aggregation and Myopic Terminal Functions</td>
<td>288</td>
</tr>
<tr>
<td>5-19</td>
<td>Results of Stage Aggregation using Pruning Set II</td>
<td>289</td>
</tr>
<tr>
<td>5-20</td>
<td>First Period Investments and Total Expected Values for Pruning Sets I through IV</td>
<td>290</td>
</tr>
<tr>
<td>6-1</td>
<td>Comparison of Sequential and Non-Sequential Solutions</td>
<td>318</td>
</tr>
<tr>
<td>6-2</td>
<td>Tradeoffs between Technology Adaptation and the Adaptive Nature of Strategies</td>
<td>322</td>
</tr>
<tr>
<td>6-3</td>
<td>Variations in Strategy Selections due to Changed Priors and Conditionals</td>
<td>325</td>
</tr>
</tbody>
</table>
APPENDIX D
AN EXTENSION TO THE LEARNING ADAPTIVE SEQUENTIAL MODEL
FOR UNKNOWN CONDITIONALS

D.1 LEARNER I: An Adaptive Model for Unknown Conditionals

D.1.1 Introduction

One of the authors has proposed a general model of complex
problem-solving process of a particular kind: "hierarchically-structured
sequential decision problems" [1]. This model is based upon Bayesian
decision theory, and requires from the engineer two kinds of probabilis-
tic inputs in a particular problem:

a) the engineer's estimate of his chances of success,
in the form of a subjective probability distribution
\[ f_j(\theta); \]
b) the engineer's estimate of the relative amount of infor-
mation gained from exercising any particular operator, \(^2\)
i, in the course of solving his problem; this is in the

---

The research reported in this appendix was originally developed with
Professor Manheim at the Universidad de Los Andes in Bogotá, Colombia,
as an extension to his Ph. D. thesis [1]. It is included here for two
reasons: (1) because of its relevance to the general problem of invest-
ments under uncertainty, and (2) most importantly, because so many
people new to the world of subjective probability ask "where does one
ger the conditionals—how can you be certain they look like that in
such an ill-defined system?" This section attempts to answer that prob-
lem by allowing yet another level of uncertainty over the conditional
distributions themselves. This allows us not only to learn about the
true underlying distribution as in Chapter 6.0, but about the condi-
tional distributions as well.

Operators, in the highway location process, are the processes which
generate and evaluate the actions. For example, location bands, bands
of interest, locations, etc., are each characterized by an operator.

D-1
form of a conditional probability distribution $g_i(y/\theta)$, where $y$ is the cost estimate produced by operator $i$.

In the model, as the engineer uses various operators, $i$, to generate new actions, $j$, and obtain cost estimates $y$, he revises his distributions $f_j(\theta)$ according to 1) the ways in which the actions $j$ are related, and 2) Bayes Theorem. This theorem is a rule for computing $f''_j(\theta/y)$, the subjective probability distribution over $\theta$ after a result $y$ has been observed, from the prior $f'_j(\theta)$ and the conditional $g_i(y/\theta)$:

$$f''_j(\theta/y) = \frac{f'_j(\theta) g_i(y/\theta)}{\sum_{\theta} f'_j(\theta) g_i(y/\theta)}$$

In this first model, the operator characteristics $g_i(y/\theta)$ are assumed known. Thus, in the process of solving a highway location problem it is assumed that these characteristics are known for each operator that may be used, and that information is acquired only about $\theta$, the cost of locations, for different parts of the terrain (different actions $j$).

An important question is the determination of these operator characteristics—for example, in the highway location problem, what are the relative degrees of information provided by the DTM location System, the DTM design System, or simpler procedures based upon tables of cross-section volumes [3]? For the capacity expansion problem presented in Chapter 6.0, how can we use information based on initial investments to revise our estimate of the learning mechanism as well as the underlying stochastic structure. The simple model described here is designed to answer this question and is implemented in a computer program, LEARNER I.

---

1 For a description of this system, see Roberts and Suhrbier [2].
D.1.2 Method

The method is a relatively straightforward extension of the earlier model. When an operator is used to produce an action and a cost estimate \( \hat{y} \) for that action, we let the information represented by that cost revise both knowledge about the terrain, represented in distributions \( f_j(\theta) \), and knowledge about which probability distributions characterize the operators, represented in distributions \( P(\beta) \), where a value of \( \beta \) corresponds to a particular set of probability distributions, one for each operator \( i \).

More precisely:

1. Let there be \( \text{MGTOT} \) probability distributions \( g_{\text{MG}}(y/\theta) \) under consideration (\( \text{MG} = 1, \ldots, \text{MGTOT} \)).

2. Let \( \text{MB} \) denote a particular value of \( \beta \); then, \( \text{MG} = \text{MBETA} (\text{MB}, \text{II}) \) identifies the probability distribution (\( \text{MG} \)) for operator \( \text{II} \) under the condition that \( \text{MB} \) is true. MBETA is a table which specifies for each operator \( \text{II} \) which distribution \( \text{MG} \) applies if \( \beta = \text{MB} \). (\( \text{MB} = 1, \ldots, \text{NBET} \)). (See Figure D-1.)

3. We assume the engineer establishes two independent priors:
   a) \( P'(\beta) = \text{PROB(MB)}, \text{MB} = 1, \ldots, \text{NBET} \). This is his estimate of the relative likelihood of different combinations of probability distributions for the operators.
   b) \( P'(\theta) = \text{THETA(JJ,MT)}, \text{where JJ denotes an action j, and MT corresponds to } \theta \). This is the prior judgement over costs of locations, \( f_j'(\theta) \).
a) \[ \text{MG} \quad \text{DISTRIBUTION} = g_{MG}(y/\theta) \]

\begin{tabular}{c|cc}
  & 1 & 2 \\
\hline
1 & \hspace{2cm} & \hspace{2cm} \\
2 & \hspace{2cm} & \hspace{2cm} \\
\end{tabular}

b) \[ \text{MG} = \text{MBETA(MB,II)}: \]

\begin{tabular}{c|cc}
  $\beta$ & II = 1 & 2 \\
\hline
MB = 1 & 1 & 1 \\
2 & 1 & 2 \\
3 & 2 & 1 \\
\end{tabular}

c) \[ \text{IMPLICATIONS OF (a) AND (b):} \]

\begin{tabular}{c}
\text{OPERATOR} \\
$\beta_1$ & $\beta_2$ & $\beta_3$ \\
\hline
II = 1 \\
\hline
= 2 \\
\end{tabular}

Figure D-1 Illustration of use of $\beta$.  

D-4
4. In line with the assumption made in the earlier model, we let \( g_{MG}(y/\theta) \) be simply a function of \( (y-\theta) \): 
\[
g_{MG}(y/\theta) = h_{MG}(\theta-y) .
\]
We measure \( \theta \) and \( y \) on the same scales; \( MT \) corresponds to \( \theta \), \( MY \) to \( y \); \( MT, MY = 1, \ldots, KDIST \). Since the range of \( (\theta-y) = (MT-MY) \) is from \((1-KDIST)\) to \((KDIST-1)\), we add a scale factor and define the variable \( KD = MT - MY + KDIST \), \( KD = 1, \ldots, KMAX \) where \( KMAX = (2*KDIST) - 1 \). Each probability distribution, indexed by \( MG \), is thus defined over \( (\theta-y) \) as a function of \( KD \): 
\[
g_{MG}(y/\theta) = DISTR(MG,KD) .
\]
Since \( MG \) is a function of \( \beta \) and \( i \), or \( MG = MBETA(\beta, II) \), we could also represent this function as \( g_i(y/\theta, \beta) \).

5. We will now derive the basic formulas for computing the posterior distributions from the priors, the operator characteristics, and the observed result \( y \).

Let \( P_j'(\beta, \theta) = P'_j(\beta) \cdot P_j'(\theta) \) be the joint prior over the unknowns \( \beta, \theta \). (We assume the priors independent.) Then,

\[
P(y, \beta, \theta) = P'(\beta, \theta) \cdot g_i(y/\beta, \theta) = P'(\beta) \cdot P'(\theta) \cdot g_i(y/\beta, \theta)
\]

\[
P(y) = \sum_{\beta} \sum_{\theta} P(y, \beta, \theta)
\]

\[
P''(\beta, \theta/y) = \frac{P(y, \beta, \theta)}{P(y)}
\]

\[
P''(\beta/y) = \sum_{\theta} P''(\beta, \theta/y)
\]

\[
= \sum_{\theta} \frac{P'(\beta) \cdot P'(\theta) \cdot g(y/\beta, \theta)}{\sum_{\beta} \sum_{\theta} P'(\beta) \cdot P'(\theta) \cdot g(y/\beta, \theta)}
\]
\[
\begin{align*}
\frac{P'(\beta) \sum_{\theta} P'(\theta) g(y/\beta, \theta)}{
\sum_{\beta} P'(\beta) \sum_{\theta} P'(\theta) g(y/\beta, \theta)}
\end{align*}
\]

and

\[
\begin{align*}
P''(\theta/y) &= \sum_{\theta} P''(\beta, \theta/y) \\
&= \frac{P'(\theta) \sum_{\beta} P'(\beta) g(y/\beta, \theta)}{
\sum_{\theta} P'(\theta) \sum_{\beta} P'(\beta) g(y/\beta, \theta)}
\end{align*}
\]

Defining

\[
\begin{align*}
g_{\beta_1}(y/\beta) &= \sum_{\theta} P'(\theta) g(y/\beta, \theta)
\end{align*}
\]

and

\[
\begin{align*}
g_{\theta_1}(y/\theta) &= \sum_{\beta} P'(\beta) g(y/\beta, \theta),
\end{align*}
\]

we have

\[
\begin{align*}
P''(\beta/y) &= \frac{P'(\beta) g_{\beta_1}(y/\beta)}{
\sum_{\beta} P'(\beta) g_{\beta_1}(y/\beta)}
\end{align*}
\]

and

\[
\begin{align*}
P''(\theta/y) &= \frac{P'(\theta) g_{\theta_1}(y/\theta)}{
\sum_{\theta} P'(\theta) g_{\theta_1}(y/\theta)}
\end{align*}
\]

This is the logic adopted in version I of the program LEARNER. (See Figure D.2)

6. The previous section (5) indicated the general outline of the computation of the posterior distributions over \( \beta, \theta \) given a result \( \hat{y} \). In the context of a hierarchically-structured problem, we generally have several actions \( j \), each with its own prior \( P_j'(\theta) \). Let \( \text{JPAR} \) denote that action to which operator \( i \) was applied to produce a new action with cost estimate \( \hat{y} \). We then follow these rules:
(10) Compute DBET(ID):
\[ g_{\theta_i}(y|\theta) = \sum_B P'(\beta) g_i(y|\beta,\theta) \quad \text{for } i = \text{IOPTR} \]
\[ \text{DBET(KD}(y-\theta)) = \sum_{MB} \text{PROB(MB)} \cdot \text{DISTR(MG(MB), KD}(y-\theta)) \]

(30) Compute DTHET(MY, MB):
\[ g_{\beta_i}(y|\beta) = \sum_{\theta} P'_j(\theta) g_i(y|\beta,\theta) \]
\[ \text{DTHET(MY, MB) = } \sum_{MT} \text{THETA(JPAR,MT)} \cdot \text{DISTR(MG(MB), KD}(y-\theta)) \]

Pick an Action, JJ

JJ = JPAR, Parent Action? Yes

JJ = 1, Universal Action? Yes

Does JJ Include Parent Action? Yes

(22) Compute Posterior over THETA(MT) for JJ:
\[ p''_j(\theta|y) = \frac{\sum_j P'_j(\theta) g_{\theta_i}(y|\theta)}{\sum_j P'_j(\theta) g_{\theta_i}(y|\theta)} \quad i = \text{IOPTR} \]
\[ j = \text{JJ} \]
\[ y = \text{MYOBS} \]

Posterior same as Prior

Have all Actions been Examined? No

Set Posterior for New Action same as Posterior over Parent

Revise: No. of Actions, NACT

Establish Inclusion List for New Action

(40) Compute Posterior over BETA(MB):
\[ p''(\beta|y) = \frac{\sum_{\beta} P'(\beta) g_{\beta_i}(y|\beta)}{\sum_{\beta} P'(\beta) g_{\beta_i}(y|\beta)} \quad i = \text{IOPTR} \]
\[ y = \text{MYOBS} \]

KNTRL=2

(Punch)
(i) the conditional distributions used are those for the particular operator, $i$, for different values of $\beta$ (obtained from the MBETA table, for each value of $\beta$;

(ii) the posterior distribution over $\beta$ is computed using the prior distribution over $\theta$ for JPAR, the action from which the new action was produced;

(iii) for all actions $j$, the posterior $P_j''(\theta)$ is

$$
P_j''(\theta/y) = \frac{P_j'(\theta) \sum_\theta g(y/\beta, \theta) P'(\beta)}{\sum_\theta P_j'(\theta) \sum_\beta P(\beta) g(y/\beta, \theta)}
$$

if $j \not\simeq$ JPAR;

$$
P_j''(\theta/y) = P_j'' \text{JPAR}(\theta/y) \text{ for } j \text{ the new action;}
$$

$$
P_j''(\theta/y) = P_j'(\theta) \text{ otherwise.}
$$

The symbol $\not\simeq$ corresponds to inclusion, a relationship among actions (represented in the program by LIST(JJ,Lj)) which gives the actions in which action JJ is included.

7. As indicated in Figure D-2, after observing result estimate $\hat{y}$ from the experiment performed, the posteriors over beta and theta are computed, then the new action (produced in the experiment executed) is added to the list by establishing its prior and the list of actions in which it is included.
D.1.3 A Comment on Initial Assumptions

In general, a large sequence of results \( y_1, y_2, \ldots \) will be input to the program LEARNER I, to get good estimates of the likelihood of values \( \beta \) and \( \theta \). The posteriors which are computed on the basis of result \( y_k \) will serve as the priors with respect to result \( y_{k+1} \).

Taking this into account, we see from section D.1.2-5 that although the priors over \( \beta \) and \( \theta \) are independent, the posteriors are not, in general:

\[
P_j''(\theta/\beta, y) = \frac{P''(\beta, \theta/\gamma)}{P''(\gamma/\beta)} = \frac{P'(y, \beta, \theta)/P''(y)}{\sum_{\theta} P'(y, \beta, \theta)/P''(y)}
\]

\[
= \frac{P'(\beta) P'(\theta) g(y/\beta, \theta)}{\sum_{\theta} P'(\beta) P'(\theta) g(y/\beta, \theta)}
\]

and

\[
P_j''(\theta/y) = \frac{P'(\theta) \sum_{\beta} P'(\beta) g_1(y/\beta, \theta)}{\sum_{\theta} P'(\theta) \sum_{\beta} P'(\beta) g(y/\beta, \theta)}
\]

from D.1.2-5,

and

\[
P''(\theta/y) \neq P''(\theta/\beta, y),
\]

or

\[
P''(\theta, \beta/y) \neq P''(\theta/y) \cdot P''(\beta/y).
\]

Therefore, using the posteriors computed as in D.1.2-5 as the priors for the next result is in error.

Note that an exact treatment of the posteriors would require us to carry, for every action \( j \), a two-dimensional prior \( P_j(\beta, \theta) \), instead of just the \( f_j(\theta) \).

Section D.2 describes the logic for this exact treatment.
D.2 LEARNER II: An Exact Formulation

D.2.1 Introduction

As described in the previous section, the use of the posteriors for experiment $e_k$ as the priors for experiment $e_{k+1}$ is in error. For a more exact treatment, we must revise the equations found in D.1.2-5. The initial assumption that $\beta$ and $\theta$ are independent and the computation for $P''(\theta/y)$ and $P''(\beta/y)$ is still valid for the first experiment. However, if we now attempt to use the same rules for obtaining $P''(\theta/y)$ and $P''(\beta/y)$ as in D.1.2-5, we violate the rules of conditional probability calculations. In other words, in the first experiment, the ability to break the joint distribution $P'(\beta,\theta)$, into the marginal distributions, $P'(\beta)$ and $P'(\theta)$, is valid. But using the same logic in the second experiment is not valid since $\beta$ and $\theta$ are now no longer independent. LEARNER II then, describes the logic behind the two-dimensional prior needed for every experiment result after the first experiment.

Note that if it is possible to show that the dependence of $\beta$ and $\theta$ after the first experiment is negligible, the more exact formulation will not be required.

D.2.2 Method

1. This, again, is an extension of the previous model, LEARNER I, and is designed to give a more exact treatment to the theory. The equations needed are now derived in a similar manner as in D.1.2-5.

$$P(y,\beta,\theta) = P'(\beta,\theta) g_1(y/\beta,\theta)$$
\[ P(y) = \sum_{\beta} \sum_{\theta} P(y, \beta, \theta) \]
\[ P_j''(\beta, \theta/y) = \frac{P_j(y, \beta, \theta)}{P_j(y)} = \frac{\sum_{\beta} \sum_{\theta} P_j(\beta, \theta) g(y/\beta, \theta)}{\sum_{\beta} \sum_{\theta} P_j(\beta, \theta) g(y/\beta, \theta)} \]

This is the two dimensional posterior used for the prior in the next experiment. The marginals are easily obtained by

\[ P_j''(\theta/y) = \sum_{\beta} P_j''(\beta, \theta/y) \]
\[ = \sum_{\beta} \frac{P_j'(\beta, \theta) g(y/\beta, \theta)}{\sum_{\beta} \sum_{\theta} P_j(\beta, \theta) g(y/\beta, \theta)} \]

and similarly

\[ P_j''(\beta/y) = \sum_{\theta} P_j''(\beta, \theta/y). \]

2. The logic of D.1.2-6 (iii), for all actions \( j \), the calculation of the posterior, \( P_j''(\theta) \), is now replaced by the above formula \( P_j''(\beta, \theta/y) \). Now \( P''(\beta/y) = \sum_j P_j(\beta/y). \)

3. The program logic is shown in Figure D-3. The only necessary addition to the variable arrays is the joint distribution over THETA and BETA, \( \text{TAB(LL,MT)} \).\(^1\) This joint distribution is two-dimensional but one is needed for every action \( j \). We therefore introduce the index LL<NBET•JJ. Thus for a problem with only 2 actions, \( JJ = 1,2 \), and \( \text{NBET} = 2 \) or \( \beta = 1,2 \), there exists two joint distributions over \( \theta \) and \( \beta \).

---

\(^1\) In the initial formulation, we were restricted to using two-dimensional arrays because of the computational device available to us at that time.
Figure D-3
Logic of the Computations for the Two-Dimensional Posterior
The joint over JJ = 1 is stored in TAB(1,MT) for all values of θ and β = 1 and TAB(2,MT) for all values of θ and β = 2. Action JJ = 2 has the joint stored in TAB(3,MT) and TAB(4,MT) for β = 1 and β = 2, respectively, and all values of θ.

4. The posterior over β,θ is computed for the first experiment's observed result, ŷ, as the product of P'(β) and P'(θ), since θ,β are assumed to be independent. The new action is added to the inclusion list and the marginal over theta calculated since this will be needed in the computation of cost estimates ŷ and the marginal over ŷ, the observed results.

D.3 Notes to Appendix D

