

CAPITAL BUDGETING UNDER RISK:  
A MATHEMATICAL-PROGRAMMING APPROACH

Alvin Keith Klevorick

Econometric Research Program  
Research Memorandum No. 89  
September 1967

Reproduction in whole or in part is permitted  
for any purpose of the United States Government.

The research described in this  
paper was supported by ONR Con-  
tract No. Nonr. 1858(16).  
NR 047-086

Princeton University  
Econometric Research Program  
Dickinson Hall  
Princeton, New Jersey

CAPITAL BUDGETING UNDER RISK:  
A MATHEMATICAL-PROGRAMMING APPROACH

Alvin Keith Klevorick

A DISSERTATION  
PRESENTED TO THE  
FACULTY OF PRINCETON UNIVERSITY  
IN CANDIDACY FOR THE DEGREE  
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE BY  
THE DEPARTMENT OF  
ECONOMICS

September, 1967

## PREFACE

It is only when one has written a dissertation that he can truly appreciate the debts of gratitude he has found expressed in the monographs of others. The debts I have incurred in writing this thesis are great. No verbal statement could adequately acknowledge them without extending the monograph unreasonably beyond its present -- perhaps, already excessive -- length.

My greatest debt of gratitude is to my thesis committee, Professor Richard E. Quandt (Chairman) and Burton G. Malkiel. Their generosity in giving of their own time, both in reading earlier versions of the manuscript and in listening to new ideas in the formative stages, is profoundly appreciated. Without their continued interest and encouragement, it is quite possible that this dissertation would not exist.

To Professor Oskar Morgenstern, Director of the Econometric Research Program at Princeton, go my sincere thanks for intellectual and financial support over the two-year period during which this project was undertaken. His constant encouragement and interest have provided an important stimulus to my research. His generous financial support of my research, including all typing costs and reproducing costs of the present manuscript, are most gratefully acknowledged. With regard to finances, I am also indebted to the National Science Foundation for providing funds in the form of a National Science Foundation Graduate Fellowship during the academic year 1965-1966 that helped support my research on this thesis.

I would also like to acknowledge my gratitude to Professor Harold W. Kuhn. It was he who introduced me to the subject of mathematical economics and to mathematical programming in particular. His help and advice with regard to the more mathematical sections of this thesis have proven most valuable.

The research in this thesis was begun during the summer of 1965 when I worked for Mathematica, of Princeton, New Jersey, on a United States Army Research Office project. To Michel L. Balinski, the director of the project, I owe a debt of gratitude for his encouragement and for my introduction to Benders' decomposition. The extent of my debt to Dr. Balinski is reflected in the numerous citations, in Chapter 8, to his excellent work. I am also indebted to Joel Cord, another member of Mathematica's staff, for his extremely careful reading of an earlier version of the material in Chapter 8. Dr. Cord's helpful comments saved the final version from several untoward remarks.

During the course of my research on the thesis, the counsel and helpful suggestions of my colleagues in the Econometric Research Program and in the Department of Economics at Princeton have been invaluable. In particular, my conversations with Charles Berry, William Branson, Michael Godfrey, Stephen Goldfeld, E. Philip Howrey, and Albert Rees have helped me to clarify my thinking at certain points and have always yielded helpful suggestions.

The speedy and efficient typing of earlier drafts of the manuscript, so important to the completion of the work, was done by Betty Kaminski, Grace Lilley, and Regina Pasche. The excellent typing of the final manuscript was the work of Manya Vas. The fine draftsmanship of Charlotte Carlson is responsible for the high quality of the figures in this work.

My last, but by no means my least, debt of gratitude is to a group of people most of whom I have never met, namely, the earlier writers on the application of mathematical programming to capital budgeting. To H. M. Weingartner, in particular, and his fellow contributors to this literature I owe the debt of stimulating my interest in the subject, a debt that will become more apparent as the reader proceeds.

While this set of acknowledgements cannot adequately convey my gratitude, it also cannot and does not remove my full responsibility for the resulting volume. Any remaining shortcomings can probably be attributed to my not having listened well enough to those whom I have just thanked for their help.



TABLE OF CONTENTS

|  | <u>Page</u> |
|--|-------------|
| Preface  | ii          |
| Table of Contents  | iv          |
| List of Figures  | vii         |
| CHAPTER ONE - INTRODUCTION   |             |
| 1.1. Introductory Remarks  | 1           |
| 1.2. The Plan of the Study   | 5           |
| CHAPTER TWO - RISK AVERSION OVER TIME  |             |
| 2.1. Introduction  | 7           |
| 2.2. Risk Aversion in the Single-Period Case   | 8           |
| 2.3. The Multiperiod Case: Framework and Problems  | 14          |
| 2.3.A. The Framework of the Multiperiod Case   | 14          |
| 2.3.B. The Multiplicity of Insurance Policies<br>and the Problems That Result  | 16          |
| 2.3.C. Introducing An Insurance Company  | 22          |
| 2.4. The Multiperiod Case: A Proposed Solution<br>and Its Implications   | 31          |
| 2.5. Decreasing Risk Aversion in the Multiperiod Case  | 40          |
| 2.5.A. The Definition of Decreasing Risk Aversion  | 40          |
| 2.5.B. Some Preliminaries to A Set of<br>Sufficient Conditions   | 43          |
| 2.5.C. Risk Balancing Over Time  | 47          |
| 2.5.D. Sufficient Conditions for Decreasing Risk<br>Aversion in the Multiperiod Case   | 51          |
| 2.6. Concluding Remarks  | 57          |
| CHAPTER THREE - THE CAPITAL-BUDGETING PROBLEM:<br>AN INTRODUCTORY VIEW   |             |
| 3.1. Introduction  | 60          |
| 3.2. The Capital-Budgeting Problem More Precisely Defined  | 60          |
| 3.3. The Imperfect Capital Market and the Planning Horizon   | 63          |
| 3.4. The Capital-Budgeting Problem As a Programming Model:<br>A Verbal Presentation  | 68          |
| 3.5. The Programming Model of Capital Budgeting<br>and Its Analogues   | 71          |
| 3.6. Risk and the Capital-Budgeting Decision   | 75          |
| 3.7. Two Disclaimers and a Final Note at the Outset  | 78          |
| CHAPTER FOUR - THE CAPITAL-BUDGETING PROBLEM UNDER CERTAINTY:<br>THE GOAL OF THE INVESTMENT PROGRAM AND PREVIOUS<br>SUGGESTIONS FOR ACHIEVING IT |             |
| 4.1. Introduction  | 80          |
| 4.2. The Goal of the Capital-Investment Program  | 81          |
| 4.2.A. The Traditional Goal and Its Problems   | 81          |
| 4.2.B. A Utility-Maximizing Approach   | 89          |
| 4.2.C. Whose Utility Is Being Maximized?   | 90          |
| 4.2.D. The Arguments of the Utility Function   | 99          |
| 4.2.E. The Objective Function of the<br>Capital-Budgeting Model  | 104         |

|  |   |     |
|--|---|-----|
| 4.3.   | An Evaluation of Previously Suggested Rules<br>for Budgeting Capital Under Certainty                                    | 113 |
| 4.3.A.   | The Subjective Approach: Urgency and<br>and Necessity   | 116 |
| 4.3.B.   | The Payback Period  | 117 |
| 4.3.C.   | The Terborgh or MAPI Formula  | 119 |
| 4.3.D.   | The Internal-Rate-of-Return Rule  | 124 |
| 4.3.E.   | The Discounted-Present-Value Rule   | 134 |
| 4.3.F.   | Previous Programming Models:<br>The Need for Further Work   | 139 |
| 4.4.   | Some Concluding Remarks   | 146 |
| CHAPTER FIVE - THE CAPITAL-BUDGETING PROBLEM UNDER CERTAINTY:<br>A NEW PROGRAMMING APPROACH                                    |   |     |
| 5.1.   | Introduction  | 147 |
| 5.2.   | The Normative Nature of the Present Model: A Review<br>of Some Findings on the State of the Art of<br>Capital Budgeting | 148 |
| 5.3.   | The Programming Model's Period-Analytic View and Its<br>Treatment of Net Returns from Investments                       | 152 |
| 5.4.   | The Physical Investment Projects and Physical<br>Interdependences   | 160 |
| 5.5.   | Cash-Flow Interactions Among Projects   | 166 |
| 5.6.   | The Capital-Market Opportunities  | 178 |
| 5.7.   | The Cash-Withdrawal Activity and the Objective Function   | 181 |
| 5.8.   | Completing the Constraint Set   | 184 |
| 5.9.   | The Complete Programming Model: A Summary   | 185 |
| CHAPTER SIX - THE CAPITAL-BUDGETING PROBLEM UNDER RISK:<br>THE RISK ENVIRONMENT AND PREVIOUS SUGGESTIONS<br>FOR COPING WITH IT |   |     |
| 6.1.   | Introduction  | 188 |
| 6.2.   | The Nature of the Risk Environment and the Need<br>for a Programming Approach   | 190 |
| 6.3.   | An Evaluation of the More Practical Previous<br>Suggestions for Budgeting Capital Under Risk                            | 195 |
| 6.3.A.   | The Finite-Horizon Approach   | 195 |
| 6.3.B.   | The Risk-Discount Approach  | 196 |
| 6.3.C.   | Sensitivity Analysis  | 200 |
| 6.4.   | An Evaluation of the More Theoretical Previous<br>Suggestions for Budgeting Capital Under Risk                          | 201 |
| 6.4.A.   | The Chance-Constrained Programming Approach   | 202 |
| 6.4.B.   | The Mean-Variance Approach: Its Rationale   | 207 |
| 6.4.C.   | The Mean-Variance Approach: Its Versions<br>and Origins   | 215 |
| 6.4.D.   | The Mean-Variance Approach: Some Objections   | 219 |
| 6.4.E.   | The Time-State-Preference Approach  | 231 |
| 6.5.   | A Closing Word  | 234 |

CHAPTER SEVEN - THE CAPITAL-BUDGETING PROBLEM UNDER RISK:  
AN AXIOMATIC APPROACH TO A PROGRAMMING MODEL

|        |  |     |
|--------|--|-----|
| 7.1.   | Introduction   | 235 |
| 7.2.   | The Expected-Utility Theorem as the Basis for Behavior Under Risk                | 236 |
| 7.3.   | The Axioms   | 238 |
| 7.4.   | A Utility Function for Capital Budgeting Under Risk                              | 250 |
| 7.4.A. | A Mathematical Statement of Sufficient Conditions for Satisfying the Axioms      | 252 |
| 7.4.B. | A Separable Cubic Utility Function that Satisfies the Axioms                     | 253 |
| 7.4.C. | A Nonseparable Cubic Utility Function that Satisfies the Axioms                  | 259 |
| 7.4.D. | A Possible Interpretation of the Proposed Utility Function                       | 274 |
| 7.5.   | The Capital-Budgeting Utility Function and Attitudes Toward Risk Characteristics | 276 |
| 7.6.   | A Programming Model for Capital Budgeting Under Risk                             | 281 |
| 7.6.A. | Some Additional Assumptions  | 281 |
| 7.6.B. | The Objective Function Expressed in Terms of the Projects                        | 285 |
| 7.7.   | A Recapitulation and A Look Ahead  | 287 |

CHAPTER EIGHT - SOLVING THE PROBLEM OF CAPITAL BUDGETING UNDER RISK

|        |   |     |
|--------|---|-----|
| 8.1.   | Introduction  | 290 |
| 8.2.   | Benders' Decomposition Applied to a Maximization Linear Mixed-Integer Programming Problem | 291 |
| 8.3.   | Benders' Decomposition and the Capital-Budgeting Model                                    | 297 |
| 8.3.A. | The Linear Subproblem   | 301 |
| 8.3.B. | The Restricted Integer Subproblem   | 307 |
| 8.3.C. | An Alternative Interpretation of the Extreme-Point Constraints                            | 316 |
| 8.4.   | Some Particular Aspects of the Solution Procedure   | 320 |
| 8.4.A. | A Justification for Examining These Extreme-Point Constraints                             | 320 |
| 8.4.B. | The Set of Special Extreme-Point Constraints  | 325 |
| 8.5.   | A Summary of the Iterative Decision Procedure   | 338 |

CHAPTER NINE - SUMMARY AND DIRECTIONS FOR FURTHER WORK

|      |                             |     |
|------|-----------------------------|-----|
| 9.1. | Summary                     | 341 |
| 9.2. | Directions for Further Work | 344 |

|              |     |
|--------------|-----|
| BIBLIOGRAPHY | 350 |
|--------------|-----|

|          |     |
|----------|-----|
| ABSTRACT | 362 |
|----------|-----|

LIST OF FIGURES

|   | <u>Page</u> |
|---|-------------|
| Figure 2.1. A's Indifference Map  | 18          |
| Figure 2.2. B's Indifference Map  | 18          |
| Figure 2.3. The Risk Averter and the Insurance Company:<br>Discounted-Present-Value Objective Function<br>for the Insurance Company   | 30          |
| Figure 2.4. An Indifference Map Between Dollars in Period $s$<br>and Dollars in Period $t_0$  | 36          |
| Figure 4.1. The Case of Divergent Borrowing and Lending Rates<br>With (1) Neither Rate the Correct Discount Rate<br>and (2) The Borrowing Rate the Appropriate<br>Discount Rate | 88          |
| Figure 4.2. Choosing Between Mutually Exclusive Alternatives  | 129         |
| Figure 7.1. The Friedman-Savage Utility Function  | 242         |
| Figure 7.2. The Markowitz Utility Function  | 242         |

## CHAPTER 1

### INTRODUCTION

#### 1.1 Introductory Remarks

Modern economists appear to have heeded well the teaching of one of their well-known precursors, Adam Smith. The degree of specialization that exists within the field of economics can serve as a fine example of the application of Smith's central principle -- the division of labor. In hope of increasing their productivity as a group, economists appear to have partitioned their field into smaller areas with subgroups of economists working in each of these more narrowly defined areas. Of course, sometimes these lines of division are cut across and many economists, with different interests, are concerned with the same problem. The capital-budgeting problem is one such issue of common concern.

Economic theorists, particularly those concerned with the theory of the firm, have shown great interest in the capital-budgeting problem. Their colleagues in macroeconomic theory have also been concerned with this problem as it affects the size and composition of the entire economy's capital stock. Public finance and economic development experts have found the question of capital allocation an important one not only in considering questions of public investment -- as in the case of water-resource development and investments of nationalized industries -- but also in designing tax incentives to stimulate private investment by firms. At the same time those economists primarily concerned with monetary theory and the theory of financial markets have seen in the capital-budgeting problem a close analogue to investors' portfolio problems in which they are interested.

Naturally, for the more business-oriented economists the capital-budgeting area is one of great interest and importance. Included in this group are the economists from business schools, particularly those concerned with corporation finance, and the students of industrial organization and public control of business. More recently, mathematical economists have also shown interest in the problem. Its relation to mathematical programming served as the force attracting this last group of students to research on the capital-budgeting decision.

This question of capital budgeting in which so many different types of economists have taken an interest is concerned with the investment-decisionmaking process of economic units, be they firms, government agencies, or other enterprises. It involves decisions about the total amounts to be invested in different periods, the different investment projects which ought to be undertaken and the times at which these ventures ought to be started, and the way in which the capital expenditures are to be financed (for example, the composition of financing as between debt and equity).

These three interrelated decisions -- the total volume of gross investment, the composition of this total investment, and the composition of the financing of these ventures -- constitute one of the major and most complex realms of entrepreneurial decisionmaking. As one student of firm investment decisions has written, "Capital budgeting represents in some respects the central problem of the firm. The complexity of the problem derives from the fact that any set of actions taken today has consequences at later times, and the opportunities available at later dates are related to decisions being implemented currently."<sup>1</sup> The complexity of the capital-budgeting problem as a whole also derives from the interdependence of the three different decisions that it encompasses. Since each decision affects the other two, a simultaneous solution really seems to be in order.

---

<sup>1</sup>Weingartner (1963, p. 139).

But, as the introduction to a collection of essays in capital budgeting tells us, "A necessary first step toward establishing a total system for these three decisions is to establish a correct procedure for arriving at each of them under the artificial assumption that the remaining two are given."<sup>2</sup> There seems to be adequate room for a further contribution to aid in taking this necessary first step. It is hoped that what follows will help in continuing this beginning. It considers the problem of selecting the investment projects to be pursued and the points in time at which they should be started, assuming the total amounts to be invested in the different periods and the composition of the financing of these investments are given.

The pervasive presence of risk or uncertainty in the environment in which this capital-budgeting decision is made has long been recognized by capital-budgeting theorists and practitioners. For example, one of the earliest modern writers in capital budgeting, who also implemented his theoretical apparatus in actual investment decisions, J. Dean, wrote that "the yardstick (for investment decisions) should permit simple adjustments to allow for ranges of uncertainty in the earnings estimates, since one of the facts to be taken into account is man's inability to see very far into the future with any great precision."<sup>3</sup> No matter what decisionmaking unit is involved, the capital-budgeting problem must be treated in a risk or uncertainty setting if it is to be treated properly.

For the most part, however, contributors to the literature on capital budgeting have confined their discussions to a world of certainty. While there are notable deviations from this well-tread path -- which will be discussed in Chapter 6 -- by and large, the literature proposing methods of capital budgeting

---

<sup>2</sup>Solomon (1959, p. 14).

<sup>3</sup>Dean (1954, p. 25).

has not adequately considered the questions of imperfect foresight, risk, and uncertainty. As W. J. Baumol writes, "Unfortunately, capital budgeting is the one subject where we can least afford to abstract from limitations in our knowledge of the future, because, by its very nature, the investment decision can only be justified in terms of its prospective effects."<sup>4</sup> He goes on to conclude the introduction to his chapter on the subject by saying "Nevertheless, we must proceed, as does most of the literature, largely without the presence of the leading character in the drama."<sup>5</sup>

One of the leading students of capital budgeting, H. M. Weingartner, took a similar position in his book on the subject. In his introductory chapter we are told, "The models proposed and examined all assume that future interest rates and future flows associated with investments are known with certainty. In making this assumption we follow the almost universal practice in the literature on capital budgeting."<sup>6</sup> In his concluding remarks, he suggests as a most desirable extension "a more explicit treatment of uncertainty, not merely as was done here, in terms of conditions imposed by the lender of funds, but also to include the outcome of investments as stochastic variables."<sup>7</sup>

The important fact is that the deterministic framework in which the capital-budgeting problem has most often been analyzed must be viewed as a first step. It has been a very worthwhile first step, but the actual decisions are necessarily made in a risky, nondeterministic setting. If theory is to come closer to analyzing capital-budgeting decisions accurately and to helping capital-budgeting practitioners, the theory must cope with this risky, nondeterministic environment.

---

<sup>4</sup>Baumol (1965, p. 434).

<sup>5</sup>Ibid., pp. 434-435.

<sup>6</sup>Weingartner (1963, p. 5).

<sup>7</sup>Ibid., p. 193.



In what follows we shall attempt to do this. The present study will constitute another of the exceptions to the general stream of capital-budgeting literature, as it attempts to carry out the extension suggested by Weingartner. The ultimate concern of this study is the formulation and solution of the problem of capital budgeting in a risk environment.

## 1.2. The Plan of the Study

The capital-budgeting problem as viewed in this study is thus a problem of decisionmaking over time in a risk environment. The next chapter begins the study with a discussion of attitudes toward risk when the planning horizon of the decision-maker extends over more than one period. In particular, the characteristics of risk aversion and decreasing risk aversion are considered in a multiperiod context. The characterization of attitudes toward risk that emerges from Chapter 2 is fundamental to the construction of the programming model of capital budgeting under risk that we will develop.

The following chapter provides an introductory view of the capital-budgeting problem. The problem, as considered in this study, is more precisely defined. The imperfect capital market in which the firm is assumed to operate is delineated and the presence of risk is briefly discussed. Chapter 3 also includes a verbal presentation of the capital-budgeting decision as a programming problem and compares the resulting model with those of other decisionmaking situations.

In Chapters 4 and 5 the problem of budgeting capital in a world of certainty is discussed as a logical first step in approaching the question of capital budgeting under risk. Chapter 4 considers the goal of the investment program and argues for a particular formulation of this objective. Procedures that have been suggested in the literature for selecting investment proposals are then evaluated in terms of their success (or lack of it) in achieving this objective. Having found these procedures, including previous programming approaches to capital budgeting,

wanting, Chapter 5 presents a new programming model of the capital-budgeting decision. Despite the shortcomings of the earlier programming models, the debt of the present author to the writers, particularly, H. M. Weingartner, who have suggested a mathematical-programming approach to capital budgeting should be quite clear. The model presented in Chapter 5 relies heavily on the work of these earlier authors whose genius it was to see the applicability of programming methods to the capital-investment decision.

Having completed the discussion of capital budgeting under certainty, Chapters 6 and 7 go on to discuss the problem of selecting the optimal set of investment projects in the presence of risk as well as imperfect capital markets. Chapter 6 sets out more precisely the nature of the risk environment. It then discusses and evaluates the suggestions that have appeared in the literature for coping with the presence of risk when a capital-budgeting decision must be made. Particular attention is given to the popular mean-variance approach to project selection. Its rationale, origins, and shortcomings are all closely examined. In Chapter 7 the programming model developed in the discussion of the certainty model is reconsidered in light of the presence of risk. Proceeding in an axiomatic fashion, the properties an objective function for capital budgeting under risk should satisfy are described, and a suitable objective function is exhibited. The objective function and the constraint set are joined together to yield a programming model of the capital-budgeting decision in a risk environment.

Chapter 8 discusses a particular solution procedure for the resulting programming model. The economic interpretation of the procedure's path to a solution of the capital-budgeting problem is presented as part of this discussion.

The study concludes with Chapter 9, which contains a summary of the development of the work, a brief discussion of the operationalism of the model presented, and indications of possible extensions of what has been done here.

## CHAPTER 2

### RISK AVERSION OVER TIME

#### 2.1. Introduction

Risk aversion (or risk preference), as usually considered in the literature, characterizes a decisionmaker's response to a single-period "fair gamble" or "actuarially neutral risk".<sup>1</sup> The experiment determining the decisionmaker's audacity or caution generally takes the following form. Suppose the individual possesses assets (or money or wealth) of value  $W$ . Confront him with a risk or gamble with expected value equal to zero, that is, one in which the expectation involves no change from his present position. If the individual prefers his status quo to accepting the gamble he is a risk averter while a preference for the gamble would show him to be a risk lover.

Numerous writers, dating back to Marshall,<sup>2</sup> have stated and discussed the relationship between an individual's attitude toward risk and the shape of his utility function. Taking the expected-utility hypothesis as the fundamental basis for behavior in a risk environment, the decisionmaker pursues that course of action which maximizes the expected value of his numerical-valued utility function  $u(W)$ . Then the individual is a risk averter if and only if  $u''(W) \leq 0$  and he is a risk lover if and only if  $u''(W) \geq 0$ . That is, risk aversion is characterized by a concave utility function while risk loving is characterized by a convex utility function.

The utility function whose shape is crucial in these definitions is a single-period utility-of-wealth function. "Strictly speaking," writes Pratt, "we are concerned with utility at a specified time (when a decision must be made) for money at a (possibly later) specified time."<sup>3</sup> But many decisions involving risk

<sup>1</sup>Friedman and Savage (1948, pp. 72-76); Markowitz (1952, pp. 151-152; 1959, pp. 215-218).

<sup>2</sup>Marshall (1920, p. 843); Arrow (1963, p. 26; 1965, p. 31); Friedman and Savage (1948, pp. 72-76); Markowitz (1952, pp. 151-152; 1959, pp. 215-218); Pratt (1964, pp. 124-127).

<sup>3</sup>Pratt (1964, p. 123).

as a significant factor are not made in a one-period or point-input (decision made) point-output (money results) framework. The investment planning of a firm or the purchase of consumer durables by a household are examples of risky decisions in which a decision taken now has a future stream of effects rather than a result at a single future (or present) point. The question naturally arises, then, as to whether the discussion of risk aversion in the single-period case can be meaningfully extended to such a multiperiod context.

The subject of this chapter is just that question -- risk aversion over time. It will be seen that attempting to extend the ideas about risk-averse (or risk-loving) utility functions or decisionmakers to the multiperiod case creates certain problems not confronted in the usual one-period discussions. Nevertheless, the concept of risk aversion can be generalized in a meaningful way to this broader case. In the discussion, special attention will be paid to the questions of increasing and decreasing absolute risk aversion upon which the recent work of Arrow and Pratt<sup>4</sup> has shed much light.

The chapter begins with a brief review of some of the results for the one-period case and some further ideas about the concept of decreasing risk aversion. Next, the framework for the multiperiod generalization is presented and the central problem this added generality creates is discussed. A solution to this resulting dilemma is then provided. The chapter continues with a discussion of the meaning of increasing and decreasing absolute risk aversion in the multiperiod case. A set of sufficient conditions for decreasing risk aversion in the multiperiod sense is presented and the chapter concludes with a proof of their sufficiency.

## 2.2. Risk Aversion in the Single-Period Case<sup>5</sup>

Following Pratt, consider a decisionmaker with assets  $W$  and (single-period) utility function  $u(W)$ . The utility function is unique up to a positive linear

<sup>4</sup>Arrow (1963, 1965); Pratt (1964).

<sup>5</sup>The review of the single-period results is largely a summary of parts of the article by Pratt cited previously, with some notational changes.

transformation, in accordance with Von Neumann-Morgenstern utility theory, and is bounded so that the expected-utility theorem may be used.<sup>6</sup> Moreover,  $u(W)$  is taken to be at least twice differentiable with a positive first derivative (positive marginal utility).

The individual is confronted with a risk or gamble that takes the form of the random variable  $\tilde{z}$ , the tilde indicating the stochastic nature of the variable. He will receive (or pay, depending on the sign of the realized value  $z$ ) an amount which depends on the realization "drawn from" the distribution of the random variable  $\tilde{z}$ . As Pratt shows, one may -- without loss of generality -- restrict attention to actuarially neutral risks, risks for which  $E(\tilde{z}) = 0$ . In all of what follows only such "fair gambles" will be considered.

The questions associated with risk aversion now center on the nature of the risk premium  $\pi$ , which depends on the level of assets and the distribution of the risk so that  $\pi = \pi(W, \tilde{z})$ . This risk premium is the deterministic sum such that the individual would be indifferent between receiving  $\pi$  dollars less than the actuarial value  $E(\tilde{z})$  -- hence, in our case receiving  $-\pi$  dollars -- and facing the risk  $\tilde{z}$ . In equation form,  $\pi$  is defined by

$$(2.1) \quad u(W-\pi) = E\{u(W+\tilde{z})\},$$

where we write  $u(W-\pi)$  instead of  $u(W+E(\tilde{z})-\pi)$  since  $E(\tilde{z}) = 0$ .

One might, alternatively, discuss the risk aversion or risk preference of a decisionmaker in terms of the amount of money he would pay an insurance company to assume his risk. This deterministic amount,  $\pi_I(W, \tilde{z})$ , his insurance premium, would be defined by

---

<sup>6</sup>Pratt does not require that the utility function be bounded. The need for boundedness of the utility function if one is to apply the expected-utility hypothesis was first observed by Menger in Menger (1934). Arrow (1963, 1965) also emphasizes the requirement that the utility function be bounded if the expected-utility result is to be used.

$$(2.1') \quad u(W - \pi_I) = E\{u(W + \tilde{z})\}$$

A comparison of (2.1) and (2.1') shows that when  $E(\tilde{z}) = 0$ ,  $\pi$  equals  $\pi_I$ . This occurs because with  $E(\tilde{z}) = 0$ , the left-hand side of (2.1), which would generally be  $u(W + E(\tilde{z}) - \pi)$ , becomes  $u(W - \pi)$ . With actuarially neutral risks, then, as all the risks discussed here are, the decisionmaker's risk premium equals his insurance premium. For reasons that will become clear in the discussion of the multiperiod case, in what follows we shall generally speak in terms of the insurance premium.

Since the decisionmaker's utility function,  $u(W)$ , is continuous with positive marginal utility, the function  $u(W - \pi)$  is a strictly decreasing, continuous function of  $\pi$  for a given  $W$ . But, then, the insurance premium or risk premium is uniquely defined by (2.1). For any given actuarially neutral risk  $\tilde{z}$  and any given level of wealth or consumption income  $W$ , there is a single amount the decisionmaker would pay to avoid facing  $\tilde{z}$ .

In terms of these new definitions, when and to what degree is an individual a risk averter? The decisionmaker is a risk averter if and only if his insurance premium is nonnegative for all  $W$  and  $\tilde{z}$ :  $\pi(W, \tilde{z}) \geq 0$ , all  $W$  and  $\tilde{z}$ . If  $\pi(W, \tilde{z}) \leq 0$  for all combinations of asset level and risk, the decisionmaker is a risk lover. Pratt then introduces a function  $r(W)$ , called the local risk aversion function, to measure the degree of aversion a utility function shows to small actuarially neutral risks. The function is defined as

$$(2.2) \quad r(W) = - \frac{u''(W)}{u'(W)}$$

Since an individual is risk-averse if and only if his utility function is concave, that is,  $u''(W) \leq 0$ , the risk premium  $\pi(W, \tilde{z}) \geq 0$  if and only if  $r(W) \geq 0$ . The magnitude of  $r(W)$  measures the extent of his aversion to risks that are actuarially neutral and small in the sense of having small variances:  $E(\tilde{z}) = 0$  and  $\sigma_{\tilde{z}}^2$  infinitesimal.

While the sign of  $u''(W)$  is thus crucial in determining whether or not a decisionmaker is a risk averter, the magnitude of that second derivative alone is not meaningful in measuring the extent to which he avoids or loves risk. The need for normalizing, dividing by  $u'(W)$ , in order to obtain a satisfactory measure results from the second derivative's dependence on the units in which utility is measured. For example, doubling  $u(W)$  would not alter the individual's behavior since Von Neumann-Morgenstern utility functions are unique only up to a positive linear transformation. In particular, doubling  $u(W)$  would not alter the individual's attitude toward risk. But it would double  $u''(W)$ , leaving the latter as an inappropriate measure of the degree of risk aversion. A proportional change in  $u(W)$  would, on the other hand, leave  $r(W)$  unchanged as it would multiply  $u''(W)$  and  $u'(W)$  by the same constant. In  $r(W)$  we have a measure that is based upon the concavity of the utility function,  $u''(W)$ , but that is also modified so as to be invariant with positive linear transformations of  $u(W)$ .

Pratt goes on to introduce the concepts of increasing and decreasing risk aversion.<sup>7</sup> A utility function is said to exhibit (strictly) decreasing risk aversion in a global sense if  $\pi(W, \tilde{Z})$  is a (strictly) decreasing function of  $W$  for all  $\tilde{Z}$ . Similarly, it is said to show (strictly) increasing risk aversion in a global sense if  $\pi(W, \tilde{Z})$  is a (strictly) increasing function of  $W$  for all  $\tilde{Z}$ . Pratt then proves that decreasing (increasing) local risk aversion  $r(W)$  is equivalent to decreasing (increasing) global risk aversion. That is, the insurance premium  $\pi(W, \tilde{Z})$  is a (strictly) decreasing function of  $W$  for all  $\tilde{Z}$  if and only if the local risk aversion function  $r(W)$  is (strictly) decreasing, and similarly with "increasing" replacing "decreasing" in both of the relevant places. What is, in Pratt's words, "nontrivial" about this theorem is that  $r(W)$  decreasing implies

---

<sup>7</sup>The same terminological convenience used by Pratt is employed in what follows. Namely, "decreasing" is used in place of the cumbersome "nonincreasing" and "increasing" is used instead of "nondecreasing".

$\pi(W, \tilde{z})$  decreasing since  $r(W)$  is a measure of risk aversion only for "small" risks. The theorem shows that if  $r(W)$  is decreasing, that is, if  $u'(W)u'''(W) \geq [u''(W)]^2$ , the insurance premium the decisionmaker would pay to protect himself against a given absolute risk  $\tilde{z}$  -- no matter what its size -- decreases as his consumption income increases.

Why be concerned with decreasing risk aversion? Because "it seems likely that many decision makers would feel they ought to pay less for insurance against a given risk the greater their assets,"<sup>8</sup> and this is possible only if their utility functions show decreasing risk aversion. Moreover, as Arrow notes, if local risk aversion is increasing, risky investment becomes an inferior good. "This result is empirically implausible" and "we must reject the hypothesis of increasing absolute risk aversion [Arrow's term for the local risk aversion function in (2.2)]. If, on the other hand, we assume decreasing absolute risk aversion, then risky investment becomes a normal good."<sup>9</sup> Invoking some armchair empiricism, Arrow writes in a later work, "The second [decreasing absolute risk aversion] certainly seems supported by everyday observation..... [I]t amounts to saying that the willingness to engage in small bets of fixed size increases with wealth, in the sense that the odds demanded diminish. If absolute risk aversion increased with wealth, it would follow that as an individual became wealthier, he would actually decrease the amount of risky assets held."<sup>10</sup> For this latter he finds little -- in fact, no -- empirical support in the way people actually behave.

One can envision situations in which the insurance premium an individual would pay to avoid a given risk would increase as his assets increased. For example, suppose a person possesses a work of art which he values greatly, and he insures against theft of it. Now if his assets increase he may well increase the

<sup>8</sup>Pratt (1964, p. 123).

<sup>9</sup>Arrow (1963, p. 26).

<sup>10</sup>Arrow (1965, p. 35).



insurance premium he would willingly pay to insure against loss of the painting or sculpture. This seems to run directly counter to the hypothesis of decreasing risk aversion stated above. Such behavior, however, can be reconciled with that hypothesis.

First, we may say that such increasingly risk-averse behavior occurs only in exceptional cases. Decreasing risk aversion is the general rule although in certain peculiar instances an individual may buy more insurance to protect against a given risk as his income rises. Such an argument cannot, however, stand without a more specific delineation of what these special cases have in common. Examining them closely -- the art owner who pays more to protect the work of art, the family man who increases the amount of life insurance he buys as his income rises -- one finds a common thread. The object or person being insured cannot adequately be described by a monetary figure. One cannot adequately value a work of art or an individual's life, for example, with a dollar sign followed by a number. Hence, in such exceptional cases it is the nonmonetary aspect of the event being insured against that causes the decisionmaker's behavior to appear inconsistent with the hypothesis of decreasing risk aversion.

Alternatively, it might be argued that the occurrences singled out as contradictions of the existence of decreasing risk aversion are not counterexamples at all for they do not meet the conditions of the definition of decreasing risk aversion. Specifically, the so-called exceptions are merely instances in which the risk is changing and hence we are not, as the conditions of that definition require, considering changes in insurance premiums for a given risk as income increases. This is most clearly seen in the life-insurance case. Neglecting the nonmonetary aspects of the event insured against -- the individual's death -- when the individual's income increases so does the risk involved. His death would result in a greater loss of income to his family. Hence, since the risk necessarily

increases along with his income, the decisionmaker's purchase of more life insurance does not contradict the decreasing-risk-aversion hypothesis.

In summary, decreasing risk aversion seems a most plausible, and if we agree with Arrow's observations a necessary, property for a utility-of-money function to possess if it is to be used in making decisions under risk. To this statement must be added the proviso that the alternatives involved in the risk decisions for which a decreasingly risk-averse  $u(W)$  is required must be adequately characterized by the monetary quantities involved.

### 2.3. The Multiperiod Case: Framework and Problems

Having completed the discussion of risk aversion and the single-period utility function, let us turn now to the more general case in which the decisionmaker's horizon extends beyond the present period. Or, more precisely, attention is now focused on the case in which a decision made at a single point in time (now) gives rise to a whole stream of monetary effects (during future periods) rather than to a single monetary effect (in the present period or in some single future period).

#### 2.3.A. The Framework of the Multiperiod Case

The decisionmaker possesses a T-period horizon and a utility function  $U = U(W)$  defined over that horizon where  $W$  is the vector  $(W_1, W_2, \dots, W_T)$  describing the sequence of his consumption incomes over the T periods.<sup>11</sup> The utility function is again unique up to a positive linear transformation and is again assumed to be bounded. It is assumed that  $U(W)$  is at least twice continuously differentiable with respect to all elements; that is,  $U_{tt} = \frac{\partial^2 U}{\partial W_t^2}$  and  $U_{st} = \frac{\partial^2 U}{\partial W_s \partial W_t}$  exist for all  $s$  and  $t$ . Moreover, the first partial derivative of  $U$  with respect to income in any period is taken to be positive everywhere. Hence,  $U_t = \frac{\partial U}{\partial W_t} > 0$  for all  $t$  and all  $W$ .

<sup>11</sup>In general, upper-case letters will denote vectors and lower-case letters with subscripts denote elements of vectors. For example,  $z_3$  is the third element of the vector  $Z$ . The vector  $W$  will be an exception: the  $t$ th element of the vector  $W$  will be denoted  $W_t$ , that is, by a subscripted upper-case  $W$ .

so that dollars in each period have positive marginal utility no matter what the time stream of the individual's consumption incomes.

One further assumption --unnecessary in the single-period case -- will be made about  $U(W)$  or rather about the relationship between incomes in different periods. It will be assumed that dollars in any period  $t$  may be treated as an ordinary commodity. In particular, we posit the same diminishing marginal rate of substitution for substitutions in every direction as one finds in the theory of consumer behavior when many commodities are involved. This is equivalent to assuming that  $U(W)$  is a quasi-concave function: for any real number  $k$ , the set of all vectors  $W$  such that  $U(W) \geq k$  is convex. (Recall that a set is convex if it contains every line segment whose end-points are in the set.)

While in the single-period case the individual was confronted with a single risk,  $\tilde{z}$ , in the present case the decisionmaker is confronted with a vector of risks (random variables)  $\tilde{Z} = (\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_T)$ , one occurring in each period. The vector  $\tilde{Z}$  is thus a  $T$ -dimensional random variable. The distribution of each random variable  $\tilde{z}_t$  represents the decisionmaker's perception of the  $t$ th risk he faces. The probability distribution may be objective or subjective in nature. But where an actual risk is specified to exist in the  $t$ th period, it is assumed that  $\tilde{z}_t$  is not constant with probability one. Once again, without loss of generality, attention is restricted to vectors of "fair risks". That is, concern in what follows is limited to vectors  $\tilde{Z}$  such that  $E(\tilde{Z}) = 0$ , or  $E(\tilde{z}_t) = 0$  for all  $t$ .

In place of the single period's single risk premium  $\pi$ , in the multiperiod case one has a vector  $\Pi$  of risk premiums, one for each period so that  $\Pi = (\pi_1, \dots, \pi_T)$ . The vector  $\Pi$  represents a vector of money amounts such that the decisionmaker would be indifferent between facing the risks  $\tilde{z}_t$  for  $t=1, \dots, T$  and receiving the deterministic amounts  $-\pi_t$  (that is,  $\pi_t$  less than the actuarial amount  $E(\tilde{z}_t) = 0$ ) for  $t=1, \dots, T$ . A vector of risk premiums must therefore satisfy the condition

$$(2.3) \quad U(W-\Pi) = E\{U(W+\tilde{Z})\}$$

Whether a particular vector does or does not satisfy (2.3) will depend on the vector of initial assets and the vector of risks, that is,  $\Pi$  is a function of  $W$  and  $\tilde{Z}$ . That some vector  $\Pi$  does exist is guaranteed by the boundedness of  $U(W)$ , which ensures that  $E\{U(W+\tilde{Z})\}$  exists and is finite, and the assumption of positive first partial derivatives and continuous second partials for  $U(W)$ . Since attention is confined to a vector of actuarially neutral risks, a vector of risk premiums is also a vector of insurance premiums, just as was true for  $\pi$  and  $\pi_I$  in the single-period case. Thus, a vector  $\Pi_I$  is a vector of period-by-period insurance payments the decisionmaker would willingly make to avoid the risk vector  $\tilde{Z}$  (that is,  $\Pi_I$  is a vector of insurance premiums), when his consumption incomes are given by  $W$ , if and only if it satisfies (2.3).

2.3.B. The Multiplicity of Insurance Policies and the Problems That Result

We come now to the crux of the difficulties with the multiperiod case. The vector  $\Pi$  is not uniquely defined by equation (2.3). In the single-period case the assumption of a continuous utility function with positive marginal utility implied that the risk premium  $\pi$  was unique. This followed because with these assumptions  $u(w-\pi)$  was a strictly decreasing continuous function of  $\pi$  ranging over all values of  $u$ . In contrast, more than one  $\Pi$ -vector can be found -- in fact, an infinite number can be found -- which satisfy (2.3) for a given  $W$  and a given  $\tilde{Z}$ . In practical terms, this means that if an individual whose initial consumption incomes over the next  $T$  periods are  $W_1, W_2, \dots, W_T$  were confronted with the set of risks  $\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_T$ , there would be many insurance policies, described by insurance premiums  $\pi_1, \pi_2, \dots, \pi_T$ , he would be willing to purchase (or sell in the case of a risk lover) to avoid the set of risks.

Why does this occur? Given his asset vector  $W$  and the risk vector  $\tilde{Z}$ , the expected utility resulting from accepting the risks can be calculated. From the assumptions about  $U(W)$ , it follows that this expected utility  $E\{U(W+\tilde{Z})\}$  exists and is finite, say it equals  $k$ . Equation (2.3) now leads us to determine all vectors  $Y$  such that  $U(Y) = k$ . Corresponding to each of these vectors is a vector of risk premiums  $\Pi(W, \tilde{Z})$  which is found by subtracting the vector  $Y$  from the initial asset vector  $W$ . But in the multiperiod case an infinite number of asset vectors generate the same utility level  $k$ . It is exactly analogous to the fact that in the case of perfectly divisible "ordinary commodities" (apples, oranges, meat, and so on -- the labels of traditional textbook indifference-map axes) an infinite number of combinations of the several commodities can yield the same utility. All asset vectors lying on the indifference surface corresponding to  $U(W) = k$  will have utility value  $k$  and corresponding to each of them will be a vector  $\Pi$  satisfying equation (2.3).

A two-period example may clarify matters. In Figures 2.1 and 2.2 the horizontal axis corresponds to dollars in period 1 and the vertical axis to dollars in period 2. The original asset position of the decisionmaker is given by the coordinates of point  $W = (W_1, W_2)$  which lies on the indifference curve corresponding to  $U(W)$  labeled  $I_0$ . Given the risk vector  $\tilde{Z} = (\tilde{z}_1, \tilde{z}_2)$  the expected utility of undertaking the risk venture,  $E\{U(W+\tilde{Z})\}$ , is evaluated. It is assumed to equal the utility level identified with indifference curve  $I_1$ . Hence, all points on  $I_1$  have the same utility as the expected utility of the gamble.

The difference between Figures 2.1 and 2.2 is that in Figure 2.1  $I_0$  is a higher indifference curve than  $I_1$  while in Figure 2.2  $I_1$  corresponds to a higher level of utility. Individual A values the certainty more than the risk while B values the risk more than his certain status quo position. In short, A is a risk averter while B is a risk lover.

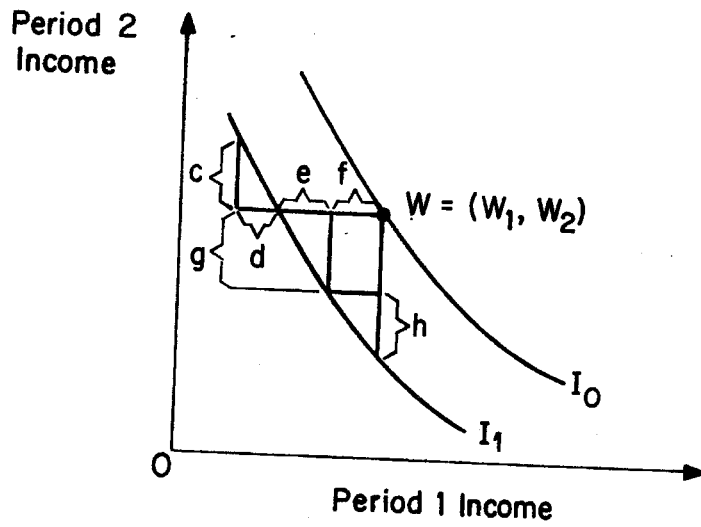


FIGURE 2.1  
A's Indifference Map

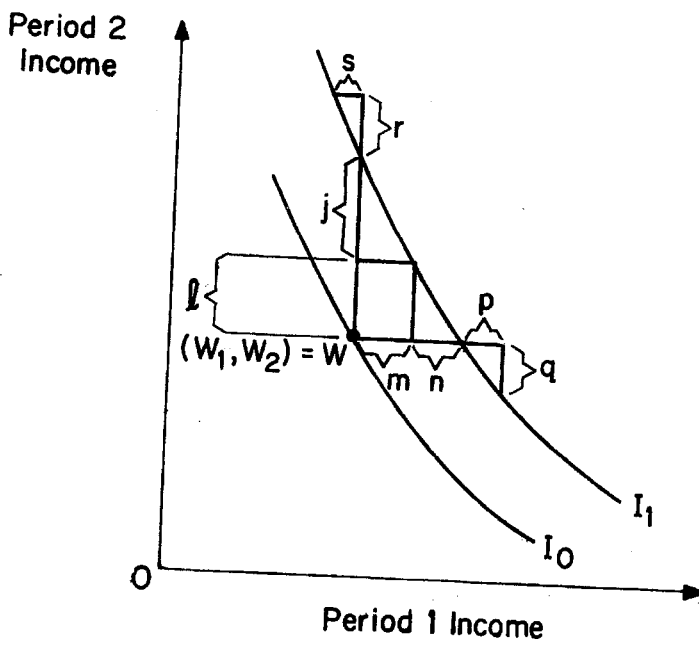


FIGURE 2.2  
B's Indifference Map

Consider any  $(y_1, y_2)$  pair on  $I_1$ . A vector of insurance premiums the decisionmaker would pay is then given by  $\Pi = (\pi_1, \pi_2) = (W_1 - y_1, W_2 - y_2) = W - Y$ . Hence, corresponding to each point on  $I_1$  there is a vector of risk premiums satisfying equation (2.3). For example, each of the following is a legitimate insurance-policy vector for individual A given the initial asset vector  $W$  and the vector of risks  $\tilde{Z}$ :  $(e+f, 0)$ ,  $(0, g+h)$ ,  $(f, g)$ , and  $(d+e+f, -c)$ . For risk-loving B sample risk-premium vectors given the same  $W$  and  $\tilde{Z}$  are  $(-m-n, 0)$ ,  $(0, -j-l)$ ,  $(-m, -l)$ , and  $(-m-n-p, q)$ .

Aside from the multiplicity of possible insurance policies, one further insight gained from the two-period example should be noted. The  $\Pi$ -vectors for the risk averter are of three different possible types, namely, (a) one positive element and one zero element, (b) one positive element and one negative element, or (c) two positive elements. In contrast, the risk lover's possible "insurance policies" include vectors with (a) one negative element and one zero element, (b) one positive element and one negative element, or (c) two negative elements. The key part of this contrast rests with parts (a). In the two-period case, the risk averter is always willing to purchase one insurance policy in which he pays the entire premium in the first period and nothing in the second,  $(e+f, 0)$  in the example in Figure 2.1, and one policy in which the whole premium is paid in the second period with no expenditure in the first,  $(0, g+h)$  in the example. On the other hand, the individual preferring risky situations is always willing to sell ( $\pi_t$  is negative) one insurance policy in which he receives the entire premium in the first period -- for example,  $(-m-n, 0)$  in Figure 2.2 -- and one policy in which the whole premium is received in the second period while he gets nothing in the first, such as  $(0, -j-l)$  in the example.

This distinction and the specific "all-in-one" policies in which the entire premium is paid (risk averter) or received (risk lover) in a single period with all

other elements of  $\Pi$  equal to zero will be returned to shortly. Such policies will form the basis for the proposed resolution of the general multiperiod problem and for the discussion of increasing and decreasing risk aversion in the multiperiod case.

For the time being, however, it is not clear how one can answer questions concerned with risk aversion -- its presence, the degree to which an individual exhibits it, and how an individual's aversion to risk changes as his vector of assets changes -- in terms of the vectors of insurance premiums and the vectors of risk premiums. In the single-period case this was simple enough since the inverse mapping

$$(2.4) \quad \pi = W - u^{-1}[E\{u(W+\tilde{z})\}]$$

was single-valued. That is, (2.4) specifies the single-period insurance premium as a function of  $W$  and  $\tilde{z}$ . In contrast, the inverse mapping

$$(2.5) \quad \Pi \in \{W - U^{-1}[E\{U(W+\tilde{Z})\}]\}$$

is a correspondence. Each vector pair  $(W, \tilde{Z})$  maps not into a single vector but into a set of vectors which shall be denoted  $\Omega(W, \tilde{Z})$ . That is, define the set  $\Omega$  by

$$(2.6) \quad \Omega(W, \tilde{Z}) = \{\Pi | U(W - \Pi) = E[U(W + \tilde{Z})]\}$$

As the problem is stated, there does not appear to be any legitimate reason for singling out a particular element of  $\Omega(W, \tilde{Z})$  and using it to answer the questions about risk aversion. Moreover, using different risk-premium vectors could provide contradictory answers to our inquiries.

To consider an example of the difficulties that one faces, return to the two-period example. It is known, by the assumption incorporated in drawing the figures, that A is a risk averter while B is a risk lover. Suppose this was unknown and one wanted to determine which of the two was more risk-averse with



regard to the  $(W, \tilde{Z})$  pair. The risk-premium vectors are to be used to answer this question. Now, if a vector of type (a) or (c) were chosen for both individuals, that is, for example,  $(e+f, 0)$  or  $(0, g+h)$  or  $(f, g)$  for individual A and  $(-m, -n, 0)$  or  $(0, -j-l)$  or  $(-m, -l)$  for individual B, it would be quite clear that A was more risk-averse. Subtracting the risk-premium vector for B from that for A would leave us with a semipositive<sup>12</sup> -- if not a positive -- vector. Decisionmaker A's risk premium is at least as large as that of B in both periods and it is greater for at least one of the two periods.

In contrast, if one chose as representative of the set of vectors  $\Omega(W, \tilde{Z})$  a  $\Pi$  of type (b) with one positive element and one negative one for both decisionmakers it is possible that one would not be able to decide which person was more risk-averse.<sup>13</sup> The difference between the two  $\Pi$ -vectors might itself

<sup>12</sup>A vector  $X = (x_1, x_2, \dots, x_T)$  is nonnegative ( $X \geq 0$ ) if  $x_t \geq 0$  for all  $t$ . It is semipositive ( $X \geq 0$ ) if  $x_t \geq 0$  for all  $t$  and  $x_t > 0$  for at least one  $t$ . Finally,  $X$  is strictly positive ( $X > 0$ ) if  $x_t > 0$  for all  $t$ .

<sup>13</sup>It might appear that applying a discount factor to the future risk premiums would enable one to decide who was more averse to risk. The individual with the larger discounted risk premium  $\pi' = \sum_{t=1}^T D_t \pi_t$  (where  $D_t$  is the discount factor to be applied to  $t$ th period dollars) would be taken to be more risk-averse. This solution is, unfortunately, illusory.

No matter whether a market rate of interest is used to form the discount factor or a subjective discount rate based on the marginal rate of substitution of dollars in the  $t$ th period for dollars in the first period is employed, discounting does not solve the problem. All asset vectors having the same utility do not have the same discounted present value using either discounting procedure. This is clear since loci of equal utility (indifference hypersurfaces) are defined by equations of the form  $U(W_1, W_2, \dots, W_T) = k$ . Loci of equal discounted present value, on the other hand, are hyperplanes defined by equations of the form  $\sum_{t=1}^T D_t W_t = k'$ . Clearly, unless  $U(W_1, W_2, \dots, W_T)$ , in fact, equals the linear utility function  $\sum_{t=1}^T D_t W_t$ , not all asset vectors of equal utility will have equal discounted present values. In particular, in the two-period case depicted in Figures 2.1 and 2.2, the loci of

have one positive element and one negative one, preventing a decision as to who is more averse to risk. But the most serious problem is that a comparison of two insurance-premium vectors of type (b) might lead to the conclusion that B was more risk-averse than A -- the direct opposite of the true situation! This could occur if the risk-premium vectors chosen for both A and B had a positive first element and a negative second element or if both had negative entries in the first location and positive ones in the second location. If, for example, one compared individual A's insurance-premium vector  $(d+e+f, -c)$  with B's risk-premium vector  $(s, -r-j-l)$ , one might find, as a result of the shapes of the utility functions involved (although admittedly not as Figures 2.1 and 2.2 are drawn),  $(d+e+f, -c) - (s, -r-j-l)$  to be a negative vector. That is, one might find the difference between A's and B's vectors to have negative elements in both positions. This would imply that B was, in fact, more risk-averse than A, which we know is not the case.

2.3.C. Introducing an Insurance Company

One possible way of resolving the problem at hand is by introducing an economic agent so far omitted from the problem's framework -- the insurance company. Perhaps "insurance exchange" would be a better name in the present context for this company plays a dual role unless all individuals are risk averters. It stands willing to sell insurance policies to risk averters -- to take on their risk vectors for stated vectors,  $\Pi$ , of risk premiums. But it is also itself willing to sell to risk lovers some of the risk vectors it has purchased from the

13 (continued) constant discounted present value would be negatively sloped straight lines. Since the indifference curves drawn in the figures are strictly convex to the origin (the utility function from which they are derived is strictly quasi-concave), at most two points on any one isoutility hypersurface would have the same discounted present value.

Hence, since the discounted present value of the initial asset vector  $W$  is given, the discounted present value of  $\Pi$ , which equals  $\sum_{t=1}^T D_t W_t - \sum_{t=1}^T D_t Y_t$ , will vary with the choice of  $Y$  and thus with the choice of the risk-premium vector. One is left in the same quandary as before. Discounting the elements of the insurance-premium vector does not help answer the relevant questions about risk aversion.

risk averters. The "company" benefits overall from its transactions with both types of individuals. In what follows, the discussion will be limited to considering the risk-averting individuals of the society. The "exchange" thus becomes an insurance company in the usual sense.

The insurance company may be thought to have existed implicitly in the one-period case. There was no need to mention it explicitly there for its role was a totally passive one. As soon as one speaks of an individual being willing to pay a certain insurance premium it is natural, however, to suppose there exists an insurance company standing ready to assume the given risk for the offered premium. If one goes a step further and makes the extremely plausible assumption that the insurance company has a positive marginal utility for money, it is clear that it will require the maximum insurance premium it can obtain to accept a given risk. That is, it will charge an individual all the traffic will bear to enable him to shed the risk he faces. The insurance process in the single-period case then takes the following simple form. An individual with assets  $W$  faces a risk  $\tilde{Z}$  which he would be willing to pay at most  $\pi$  dollars to avoid. The insurance company, standing ready to accept any monetary risk for the maximum insurance premium, accepts the risk and the payment of  $\pi$  dollars.

The multiperiod case is not as simple since the insurance company itself has a multiperiod utility function it uses to evaluate any time stream of dollar receipts. Its role is no longer a passive one. The company remains willing to insure any individual against a vector of risks but it reserves the right to choose from among all  $\Pi$ -vectors of insurance premiums he offers the one it prefers. While the individual is indifferent among all  $\Pi$ -vectors in  $\Omega(W, \tilde{Z})$  the insurance company is not. As in the single-period case where it chose the insurance premium it preferred -- namely, the largest one it could exact -- so in the multiperiod case the insurance company selects from the set  $\Omega(W, \tilde{Z})$  the

vector of insurance premiums it prefers most. It selects the risk-premium vector that effects maximization of the expected value of its utility function.

Two aspects of the insurance company's optimization ought to be emphasized. First, it is the expected value of its multiperiod utility function that the company maximizes. When it receives the vector  $\Pi$  of insurance premiums from the individual, it does so only on the condition that it also accepts the vector of risks,  $\tilde{Z}$ , he faced. The amount of money the insurance company receives in each of the  $T$  periods due to a particular individual is not deterministic. Instead, it is the stochastic amount  $\pi_t^t + \tilde{z}_t$ . Faced with a problem in decision-making under risk, then, the insurance company maximizes its expected utility.

Second, in facing this risk situation, the insurance company has an advantage the individual decisionmaker does not. Recall that the insurance company is serving many individuals. Each period, it is receiving insurance premiums from many people and accepting the risks the individuals wanted to avoid. The insurance company can pool these many risk vectors and, hence, effectively lighten the burden of any particular risk vector that it must face. In its ability to diversify by accepting a number of risk vectors lies the insurance company's strength. Of course, in deciding on the appropriate insurance policy for each individual, the company must take account of the harmful interplay (for example, positive covariances) between the risks contributed by different individuals as well as the beneficial interactions (for example, negative covariances) between them. In short, the insurance company must settle upon the insurance policies it will sell to different individuals simultaneously.

In order to state the insurance company's optimization problem, let  $U_I$  denote the insurance company's utility function and let the pre-subscript  $i$  denote the  $i$ th individual. With this notation,  $U_i$  denotes the  $i$ th individual's utility function,  $W_i$  his initial asset vector,  $\tilde{Z}_i = \{\tilde{z}_t\}$  the vector of risks he

faces, and  ${}_i\Pi = \{ {}_i\pi_t \}$  the vector of insurance premiums he will pay the insurance company to relieve him of his risks. The insurance company wants to determine the  $n$  optimal insurance policies it should sell, one to each decisionmaker. These  $n$  policies emerge as the solution to the following constrained optimization problem.

$$(2.7) \left\{ \begin{array}{l} \text{Maximize } E\{U\} = E\left\{U\left(\sum_{i=1}^n [{}_i\pi_1 + {}_i\tilde{Z}_1], \dots, \sum_{i=1}^n [{}_i\pi_T + {}_i\tilde{Z}_T]\right)\right\} . \\ \text{Subject to: } {}_iU({}_iW - {}_i\Pi) = E\{ {}_iU({}_iW + {}_i\tilde{Z}) \} \quad \text{for } i=1, \dots, n . \end{array} \right.$$

It is a problem in  $n \cdot T$  variables, the  ${}_i\pi_t$ 's, and  $n$  constraints.

Each constraint in the insurance company's problem contains, and hence restricts, the values of only  $T$  variables, namely, the  $T$  risk premiums of the individual whose position the constraint represents. No variable subject to the insurance company's control, no  ${}_i\pi_t$ , appears in more than one of the  $n$  constraints. Hence, the constraints of the problem cannot be inconsistent with one another. In addition, the assumed boundedness of each  ${}_iU({}_iW)$  ensures that for each  $i$ ,  $E\{ {}_iU({}_iW + {}_i\tilde{Z}) \}$  exists and is finite, and thus ensures that there exists some  ${}_i\Pi$  satisfying the  $i$ th constraint. The absence of inconsistencies among the constraints and the existence of at least one  ${}_i\Pi$  satisfying the  $i$ th one of them, for each  $i=1, \dots, n$ , means that the insurance company's problem is always feasible.

The assumptions made earlier about the shape of each individual's multiperiod utility function,  ${}_iU({}_iW)$ , provide even more information about the nature of the insurance company's problem. Recall that each  ${}_iU({}_iW)$  is quasi-concave in  ${}_iW$ . But, then for a fixed  ${}_iW$ -vector,  ${}_iU({}_iW - {}_i\Pi)$  is quasi-concave in  ${}_i\Pi$ . In particular, with  ${}_iW$  fixed as it is for each  $i$  in (2.7), the set of  ${}_i\Pi$ -vectors such that  ${}_iU({}_iW - {}_i\Pi) \geq E\{ {}_iU({}_iW + {}_i\tilde{Z}) \}$  is convex, recalling that with  ${}_iW$  and  ${}_i\tilde{Z}$  given  $E\{ {}_iU({}_iW + {}_i\tilde{Z}) \}$  is a constant. The  $i$ th constraint in (2.7) is thus the boundary

of a convex set of  ${}_i\Pi$ -vectors.<sup>14</sup> The insurance policies that constitute the boundary with which the  $i$ th decisionmaker confronts the company are policies,  ${}_i\Pi$ -vectors, such that insofar as the individual is concerned no premium could be increased without a concomitant decrease in another. That is, for any fixed set of T-1 premiums,  ${}_i^\pi_1, {}_i^\pi_2, \dots, {}_i^\pi_{t-1}, {}_i^\pi_{t+1}, \dots, {}_i^\pi_T$ , the boundary the  $i$ th constraint in (2.7) represents shows the largest premium,  ${}_i^\pi_t$ , the insurance company could exact from the  $i$ th decisionmaker in period t.

One could in fact, consider the set of feasible insurance policies for the  $i$ th individual to be the entire set of  ${}_i\Pi$ -vectors such that  ${}_iU({}_iW - {}_i\Pi) \geq E\{{}_iU({}_iW + {}_i\tilde{Z})\}$ . But since each  ${}_i^\pi_t$  has positive marginal utility for the insurance company, it will proceed immediately to the boundary. Hence, the constraint is written as an equation in (2.7). The situation is exactly analogous to the usual consumer-theory presentation in which, although all commodity bundles on or below the consumer's budget hyperplane are feasible, the consumer is assumed to move immediately to his budget hyperplane in search of the optimal commodity bundle.

The feasible region for the insurance company is thus the set of  $nT$ -element vectors  $\{ {}_1^\pi_1, {}_1^\pi_2, \dots, {}_1^\pi_T, {}_2^\pi_1, {}_2^\pi_2, \dots, {}_2^\pi_T, \dots, {}_n^\pi_1, {}_n^\pi_2, \dots, {}_n^\pi_T \}$  satisfying the  $n$  independent constraints  ${}_iU({}_iW - {}_i\Pi) \geq E\{{}_iU({}_iW + {}_i\tilde{Z})\}$  for  $i=1, \dots, n$ . But, as just indicated, each of these constraints defines a closed convex set. Since the intersection of closed convex sets is itself a closed convex set, the feasible region of the insurance company's problem is a closed convex set. The insurance company moves immediately to the boundary of this set as the first step in its optimization process.

It should be noted that the boundary of the feasible region defined by the constraint set in (2.7) is "concave to the origin" of the insurance-premium space. This follows from the quasi-concavity of each  ${}_iU({}_iW)$  function and the negative marginal utility of  ${}_i^\pi_t$  for the  $i$ th individual. As  ${}_i^\pi_t$  increases, the consumption income the individual has in the  $t$ th period decreases. Since consumption income in

<sup>14</sup>An examination of Figure 2.1 should convince the reader that this set is convex in the two-period case presented there.

any period  $t$  possesses positive marginal utility for him,  ${}_i\pi_t$  must have negative marginal utility for the  $i$ th decisionmaker. But the iso-level surfaces of a quasi-concave function with negative first partial derivatives are concave to the origin. They show an increasing marginal rate of transformation between the quantities on any two axes. In Figure 2.1, for example, the insurance-policy possibility frontiers generated by decisionmaker A, the indifference curves lying below  $I_0$  (for example,  $I_1$ ), are clearly concave to the origin of insurance-policy space  $W = (W_1, W_2)$ . Since each of the constraints in (2.7) is concave to the origin of the insurance company's premium space and since the feasible region defined by the intersection of the  $n$  independent constraints  ${}_iU({}_iW - {}_i\Pi({}_iW, {}_i\tilde{Z})) \geq E\{{}_iU({}_iW + {}_i\tilde{Z})\}$  is a closed convex set, the boundary of that feasible region will be concave to the origin of the insurance-premium space.

If it is now assumed that  $E\{{}_I U\}$  is a quasi-concave function,<sup>15</sup> the problem in (2.7) calls for maximization of a quasi-concave function over a nonempty closed convex set that is bounded from above. The result of that optimization process will be a unique set of policies  ${}_1\Pi, {}_2\Pi, \dots, {}_n\Pi$ , one for each decisionmaker, if either  $E\{{}_I U\}$  is strictly quasi-concave or if the boundary of the constraint set is strictly concave to the origin or both. A unique set of policies emerges, then, if the constraint set's boundary contains no linear segments or if the expected-utility indifference hypersurfaces contain no linear segments or if none of these--the boundary of the constraint set or any of the indifference hypersurfaces -- contains any linear segment. This would result, for example, if each  ${}_iU({}_iW)$  were assumed to exhibit a strictly diminishing marginal rate of substitutability for any substitutions among consumption incomes in different periods. Then, every indifference hypersurface of every individual would be strictly convex to the origin, containing no linear segments. As a result, the boundary of the insurance company's constraint set would be strictly concave to the origin of its insurance-policy space.

<sup>15</sup>Note well that it does not suffice to assume  ${}_I U$  is quasi-concave because  ${}_I U$  may be quasi-concave and yet  $E\{{}_I U\}$  not quasi-concave.

With the assumptions of quasi-concavity for the insurance company's expected-utility function and strict quasi-concavity for each decisionmaker's utility function, the insurance process takes the following form. Individuals with time streams of consumption incomes given by  ${}_iW$  for  $i=1, \dots, n$ , are confronted with sets of risks, one in each period for each decisionmaker, described by  ${}_i\tilde{Z}$ , again with  $i=1, \dots, n$ . They would rather not face these risks and want to purchase insurance policies that will leave the insurance company with the risks. In exchange, each decisionmaker will make a set of certain payments  ${}_i\Pi$  to the company. After each individual specifies the set of boundary policies from which he would willingly choose, namely,  ${}_i\Omega({}_iW, {}_i\tilde{Z})$ , the insurance company chooses those policies -- one for each decisionmaker -- that maximize its expected utility. These are now the uniquely defined risk-premium vectors.

As a specific example, suppose the insurance company's utility function is a discounted-present-value one. In later chapters, an argument will be put forth against the use of such a function for decisionmaking under risk. Nevertheless, suppose the insurance company of the present example has settled upon such a utility function. Its objective function is then to maximize

$$(2.8) \quad E\{ {}_I U \} = E\left\{ \sum_{t=1}^T D_t \left[ \sum_{i=1}^n ({}_i\pi_t + {}_i\tilde{Z}_t) \right] \right\},$$

where  $D_t$  is the discount factor applied to dollars in period  $t$  to put them on an equal footing with dollars in period 1. Each  $D_t$  is a constant, and each  ${}_i\pi_t$  is a nonstochastic quantity to be determined by the insurance company. In addition, recall that consideration has been restricted to actuarially neutral risks so that  $E({}_i\tilde{Z}_t) = 0$  for all  $i$  and  $t$ . The insurance company's goal is then to maximize

$$\sum_{t=1}^T D_t \left( \sum_{i=1}^n {}_i\pi_t \right) \text{ which equals } \sum_{i=1}^n \left[ \sum_{t=1}^T D_t {}_i\pi_t \right] \text{ since the order of summation is}$$

irrelevant. In this case, then, the insurance company's problem is the following.



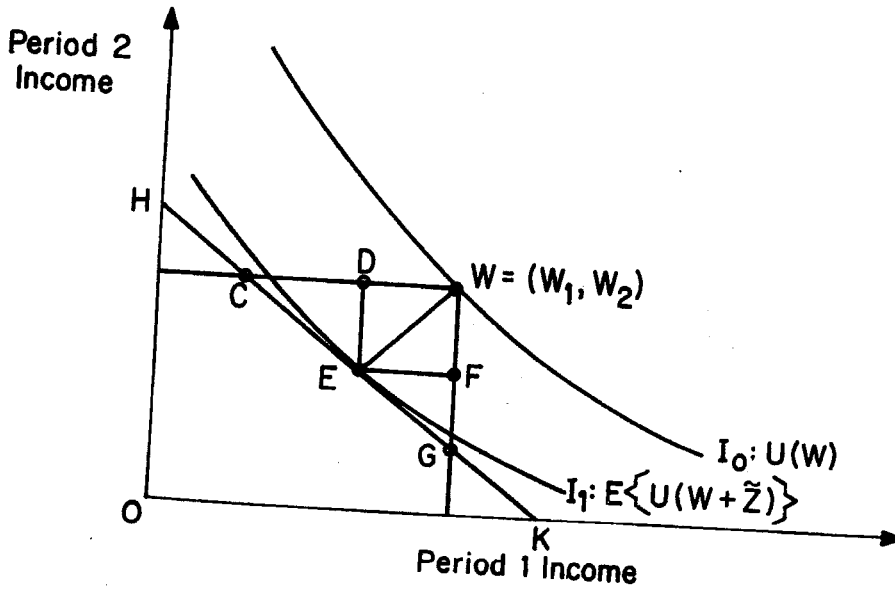
$$(2.9) \left\{ \begin{array}{l} \text{Maximize} \quad E\{U\} = \sum_{i=1}^n \left[ \sum_{t=1}^T D_t i \pi_t \right] \\ \text{Subject to:} \quad U(iW - i\Pi) = E\{U(iW + i\tilde{Z})\} \quad \text{for } i=1, \dots, n. \end{array} \right.$$

In short, the insurance company maximizes the sum of the discounted present values of all the policies it sells ( $n$  in number) subject to the  $n$  constraints imposed by the individual decisionmakers. Since the contribution of any one decisionmaker's policy to the company's objective function is independent of any other's and since the constraint each decisionmaker imposes is independent of the constraints the others impose, it is clear that the company can proceed by solving the following  $n$  problems.

$$(2.10) \left\{ \begin{array}{l} \text{Maximize} \quad \sum_{t=1}^T D_t i \pi_t \\ \text{Subject to:} \quad U(iW - i\Pi) = E\{U(iW + i\tilde{Z})\} \quad , \end{array} \right.$$

one problem for each  $i=1, \dots, n$ . Each of these problems entails the maximization of a linear function of  $T$  variables along the boundary of a convex set in those variables.

Consider, for example, the two-period case shown in Figure 2.3, where the subscript  $i$  is dropped for convenience. The individual confronted with the risk vector  $\tilde{Z}$  when his assets are  $W = (W_1, W_2)$  is indifferent among all risk-premium vectors obtained by subtracting a vector represented on curve  $I_1$  from  $W = (W_1, W_2)$ . (Indifference curve  $I_0$  again corresponds to the utility of the original asset vector while  $I_1$  corresponds to the expected utility of the risky situation.) The insurance company's utility function denoted  $U$  is, by assumption, a discounted-present-value one. Hence, in its contract with the individual whose position is shown in Figure 2.3, the insurance company desires to maximize



**FIGURE 2.3**  
**The Risk Averter and the Insurance Company:**  
**Discounted-Present-Value Objective Function**  
**for the Insurance Company**

$$(2.11) \quad \pi_1 + D_2 \pi_2, \quad ,$$

the individual's contribution to the insurance company's expected utility. It maximizes (2.11), where  $D_2$  is the discount rate applied to second-period dollars, subject to the constraint that  $U(W_1 - \pi_1, W_2 - \pi_2) = E\{U(W_1 + \tilde{z}_1, W_2 + \tilde{z}_2)\}$  for the individual decisionmaker.

Now, if  $D_2 = \frac{\overline{WC}}{\overline{WG}} = \frac{\overline{OK}}{\overline{OH}}$ , the insurance company would maximize (2.11) by

choosing from among all points on  $I_1$  the point E to define the insurance policy the risk averter buys. From among all the insurance policies the risk averter would be willing to purchase he is, in fact, offered only the policy  $(\pi_1, \pi_2) = (\overline{WD}, \overline{WF})$ . This would be the relevant  $\Pi$ -vector to use in answering questions about this decisionmaker's aversion to risk. It plays the same role in the multiperiod case as the single number  $\pi$  does in the one-period case.

It is interesting that the insurance company whose role is so totally passive in the one-period or single-stage case becomes such an important character in the multiperiod situation. The question naturally arises, though, whether anything can be said about an individual's behavior toward risk without explicitly introducing the specific utility function of the insurance company, the utility functions of all other individuals, and the asset and risk vectors of all those other individuals. In the next section it will be shown how an affirmative answer can be given to this last question.

#### 2.4. The Multiperiod Case: A Proposed Solution and Its Implications

An individual is defined to be risk-averse if he prefers (or is indifferent between) his certain status quo position  $W$  to (and) the fair risky result determined by realization of the vector random variable  $W + \tilde{Z}$ . For a risk averter and only for a risk averter,  $U(W) \geq E\{U(W + \tilde{Z})\}$  for any risk vector with  $E(\tilde{Z}) = 0$ . For a strict

risk averter the weak inequality is replaced by  $U(W) > E\{U(W+\tilde{Z})\}$ . By extending Jensen's Inequality to functions of vectors of variables, it can be shown that an individual is a risk averter if and only if his multiperiod utility function is concave. Jensen's Inequality in the single-variable case states that if  $f(x)$  is a concave (convex) function of  $x$ , then if  $x$  is a random variable  $E\{f(x)\} \leq f(E[x])$  ( $E\{f(x)\} \geq f(E[x])$ ).<sup>16</sup> The extension of the inequality to functions of many variables is easily accomplished by the following lemma.

LEMMA 2.1. If and only if a function  $f(X)$  is concave in  $X = (x_1, x_2, \dots, x_n)$  and has at least continuous first partial derivatives, then  $E\{f(X)\} \leq f(E[X])$ .  
If and only if a function  $f(X)$  is convex in  $X = (x_1, x_2, \dots, x_n)$  and has at least continuous first partial derivatives, then  $E\{f(X)\} \geq f(E[X])$ .

PROOF: (i) Concavity of  $f(X) \rightarrow E\{f(X)\} \leq f(E[X])$ .

If and only if  $f(X)$  is concave in  $X$  and has at least continuous first partial derivatives,<sup>17</sup>

$$(2.12) \quad f(X) \leq \nabla f(X^0) \cdot (X - X^0) + f(X^0),$$

for all  $X$ , where  $X^0$  is any particular vector and

$$\nabla f(X^0) = \left( \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right)_{x_1=x_1^0, x_2=x_2^0, \dots, x_n=x_n^0}.$$

Verbally, if and only if

$f(X)$  is concave, the function always lies on or below its tangent hyperplanes.

If  $X$  is a vector of random variables and  $E(X)$  the vector of means, let  $X^0 = E(X)$  in (2.12) and take expectations on both sides of the inequality. The result is

$$E\{f(X)\} \leq E\{ \nabla f(E[X]) \cdot (X - E[X]) + f(E[X]) \} \quad \text{or}$$

$$(2.13) \quad E\{f(X)\} \leq f(E[X]),$$

<sup>16</sup> See Feller (1966, pp. 151-152) for a discussion of Jensen's Inequality in the case of functions of a single variable.

<sup>17</sup> See, for example, Hadley (1964, p. 86).

since  $\nabla f(E[X])$  is a vector of constants,  $E\{X-E[X]\} = 0$ , and  $E\{f(E[X])\} = f(E[X])$ .

(ii) Convexity of  $f(X) \rightarrow E\{f(X)\} \geq f(E[X])$ .

The proof is exactly analogous to that of part (i) starting from the inequality

$$(2.14) \quad f(X) \geq \nabla f(X^0) \cdot (X-X^0) + f(X^0) ,$$

which is true if and only if  $f(X)$  is convex in  $X$  and has continuous first partial derivatives. One obtains

$$(2.15) \quad E\{f(X)\} \geq f(E[X]) .$$

(iii)  $E\{f(X)\} \leq f(E[X]) \Rightarrow f(X)$  is concave.

Proving this implication is clearly equivalent to proving that if  $f(X)$  is not concave then  $E\{f(X)\} \neq f(E[X])$ . But if  $f(X)$  is not concave, which means, in particular, it is not a linear function since a linear function is both concave and convex, there is some region of  $n$ -dimensional space over which it must be strictly convex. The strict inequality in (2.14) then applies, and it follows from part (ii) that for some region of the space of  $X$ -vectors,  $E\{f(X)\} > f(E[X])$ .

(iv)  $E\{f(X)\} \geq f(E[X]) \rightarrow f(X)$  is convex.

The proof is exactly analogous to that of part (iii), with the strict inequality in (2.12) and the development in part (i) being used this time.

The asserted equivalence of risk aversion and concavity of the multiperiod utility function follows immediately from Lemma 2.1. Let  $U(W)$  take the place of  $f(X)$  in the lemma and recall that  $E\{\tilde{Z}\} = 0$  for the fair risks used to determine risk aversion or risk preference. It then follows immediately from the lemma that  $U(W) \geq E\{U(W+\tilde{Z})\}$  if and only if  $U(W)$  is concave and  $U(W) \leq E\{U(W+\tilde{Z})\}$  if and only if  $U(W)$  is convex. (Note that  $f(E[X])$  in the lemma becomes  $U(W)$  in this special case since  $W$  is deterministic and  $E\{\tilde{Z}\} = 0$  so that  $E\{W+\tilde{Z}\} = W$ .) An individual is thus a risk averter if and only if his multiperiod utility function  $U(W)$  is concave while he is a risk lover if and only if  $U(W)$  is convex.

But can one determine whether an individual is a risk averter solely by observing "insurance policies" he would be willing to buy? The answer is yes. To do so one must, however, examine a specific subset of such policies. Consider a risk-premium vector feasible for the  $(W, \tilde{Z})$  pair and in which the individual pays nothing in any period but the  $t$ th one. That is, define the risk-premium vector  $\Pi^t(W, \tilde{Z})$  such that

$$(2.16) \quad \Pi^t(W, \tilde{Z}) \in \Omega(W, \tilde{Z}) \quad \text{and} \quad \pi_i^t = 0 \quad \text{for all } i \text{ except } i=t .$$

Since there are  $T$  periods in the planning horizon, there are  $T$  such "all-in-one-period" policies. There is an insurance policy the terms of which state that a premium is paid only in the first period, a policy which requires a premium only in the second period, and so on. These are the  $\Pi$ -vectors of type (a) as they were classified in the two-period example presented earlier. The following theorem demonstrates the importance of these "all-in-one-period" policies in assessing a decisionmaker's attitudes toward risk.

THEOREM 2.1. A decisionmaker is a strict risk averter with respect to the risk vector  $\tilde{Z}$  when his asset vector is  $W$  if and only if  $\Pi^1(W, \tilde{Z})$  is a semipositive vector, that is, if and only if  $\pi_1^1(W, \tilde{Z}) > 0$  while  $\pi_t^1(W, \tilde{Z}) = 0$  for all  $t=2, \dots, T$ .

The theorem states that an individual is a strict risk averter if and only if he would buy an insurance policy in which the entire premium must be paid in the first period. A better understanding of the theorem is achieved if the following lemma is proved first.

LEMMA 2.2. Either all  $\Pi^t(W, \tilde{Z})$  vectors are semipositive or all  $\Pi^t(W, \tilde{Z})$  vectors are seminegative. That is, either  $\pi_t^t(W, \tilde{Z}) > 0$  for all  $t$  or

$$\pi_t^t(W, \tilde{Z}) < 0 \text{ for all } t. \text{ }^{18}$$

PROOF: Suppose all  $\pi_t^t$  are positive except for one, say  $\pi_s^s$ . Select any  $t \neq s$ , call it  $t_0$ , and consider the indifference curves between dollars in period  $t_0$  and dollars in period  $s$ , holding all other elements of the vector  $W$  constant. That is, consider ceteris paribus indifference curves between dollars in  $s$  and dollars in  $t_0$ . The indifference map is shown in Figure 2.4 where  $I_0$  shows the utility level yielded by the initial asset vector  $W$ . Since  $\pi_{t_0}^{t_0}$  is assumed to be positive,  $(W_s, W_{t_0}) - (0, \pi_{t_0}^{t_0})$  lies directly below the original point  $(W_s, W_{t_0})$ . The point  $(W_s, W_{t_0} - \pi_{t_0}^{t_0})$  is labeled  $J$  in the figure. On the other hand, since  $\pi_s^s$  is assumed to be negative,  $(W_s, W_{t_0}) - (\pi_s^s, 0) = (W_s - \pi_s^s, W_{t_0})$  labeled  $H$  lies directly to the right of the original point. But since both  $\Pi^s$  and  $\Pi^{t_0}$  are elements of the set  $\Omega$ ,

$$(2.17) \quad U(W - \Pi^s) = U(W - \Pi^{t_0}) = E\{U(W + \tilde{Z})\} .$$

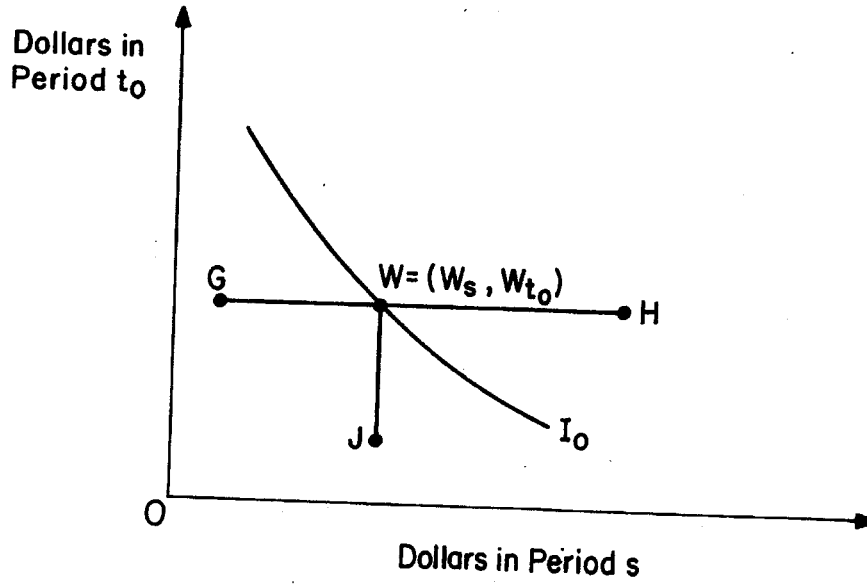
Hence an indifference curve must pass through the points  $J$  and  $H$ . But this is impossible. It contradicts the assumption that dollars in different periods are ordinary commodities with respect to one another since the latter implies that their indifference curves do not intersect. If an indifference curve had to pass through  $J$  and  $H$ , however, it would clearly have to intersect the indifference curve  $I_0$ .

Alternatively, the passing of an indifference curve through  $J$  and  $H$  runs counter to the assumption that dollars in any period have positive marginal utility given any asset vector. It does so since if  $J$  and  $H$  lie on a single indifference curve that curve must have a positive slope. But then one would have

$$(2.18) \quad \frac{dW_{t_0}}{dW_s} = - \frac{U_s}{U_{t_0}} > 0$$

---

<sup>18</sup>Where no confusion is possible, the specification of the asset-risk pair  $(W, \tilde{Z})$  is suppressed in the interests of notational simplicity.



**FIGURE 2.4**  
An Indifference Map between  
Dollars in Period  $s$  and Dollars in Period  $t_0$



which implies that not both  $U_{t_0}$  and  $U_s$  are positive, contradicting the assumption made earlier.

Hence, all  $\pi_t^t$  must have the same sign. If, for example,  $\pi_s^s$  were also positive a point as G would appear in Figure 2.4 instead of H. The indifference curve could be negatively sloped, dollars in both periods could have positive marginal utility, there would be diminishing marginal substitutability and hence no contradictions would be implied.

Clearly, the proof also holds if one assumes that all but one of the  $\pi_t^t$ 's are negative. Then it can be shown that the remaining one must also be negative.

We return now to the proof of Theorem 2.1, proving the necessity first.

PROOF: (i) Decisionmaker is a strict risk averter  $\rightarrow \pi_1^1 > 0$ .

Since he is a strict risk averter,  $U(W) > E\{U(W+\tilde{Z})\}$ . By assumption, the marginal utility of dollars in any period is positive. Hence, if all elements of  $W$  except one are held constant, then if and only if the remaining element is decreased,  $U(W)$  will decrease. Moreover, by the assumption that  $U(W)$  is at least twice continuously differentiable, by decreasing that single element sufficiently, holding all other asset levels constant, a new asset vector  $W^* \leq W$  will result such that  $U(W^*) = E\{U(W+\tilde{Z})\}$ . Since this is true for changes in any single element of  $W$ , it is true for changes in the first element only.

But by definition,  $W - \Pi^1$  is a vector differing from  $W$  in only one element, namely, the first, and  $U(W - \Pi^1) = E\{U(W+\tilde{Z})\}$ . Since  $U(W) > E\{U(W+\tilde{Z})\}$ , we have that  $U(W - \Pi^1) < U(W)$ . The preceding statements therefore imply that  $W - \Pi^1 \leq W$  and hence  $\Pi^1 \geq 0$ . Since all elements of  $\Pi^1$  are zero except for the first, it follows that  $\pi_1^1 > 0$ .

(ii)  $\pi_1^1 > 0 \rightarrow$  decisionmaker is a strict risk averter.

If  $\pi_1^1 > 0$ , then  $\Pi^1 \geq 0$  and  $W - \Pi^1 \leq W$ . In fact  $W - \Pi^1$  is identical, element by element, with  $W$  except for the first element. The first element of  $W - \Pi^1$  is

$W_1 - \pi_1^1 < W_1$ . But since it was assumed that dollars in each period always have positive marginal utility, the fact that  $W_1 - \pi_1^1 < W_1$  while  $W_t - \pi_t^1 = W_t$  for all  $t \geq 2$  means  $U(W - \Pi^1) < U(W)$ . By the definition in (2.16), however,  $U(W - \Pi^1) = E\{U(W + \tilde{Z})\}$ . Thus we have  $E\{U(W + \tilde{Z})\} < U(W)$  and the individual is a strict risk averter.

COROLLARY 2.1. A decisionmaker is a strict risk averter with respect to the risk vector  $\tilde{Z}$  when his asset vector is  $W$  if and only if  $\Pi^t(W, \tilde{Z})$  is a semi-positive vector for any  $t$ .

PROOF: This follows simply from Lemma 2.2 and Theorem 2.1. The lemma implies that  $\Pi^1(W, \tilde{Z})$  can be replaced by any other  $\Pi^t(W, \tilde{Z})$  in the statement of the theorem. If one  $\Pi^t(W, \tilde{Z})$  is semipositive, then they all are semipositive.

Another interesting fact follows from the lemma and the theorem.

COROLLARY 2.2. For any given asset vector and risk vector, a strict risk averter will always be willing to purchase at least one "boundary" insurance policy for which he pays a positive premium in each period. That is, for an individual who is a strict risk averter with respect to  $(W, \tilde{Z})$  there exists at least one vector  $\Pi(W, \tilde{Z}) > 0$  such that  $U(W - \Pi(W, \tilde{Z})) = E\{U(W + \tilde{Z})\}$ .

PROOF: For the strict risk averter  $\Pi^t(W, \tilde{Z}) \geq 0$  for all  $t$ ;  $\pi_t^t > 0$  for all  $t$ . Moreover, the vectors  $W - \Pi^t$ ,  $t=1, \dots, T$ , all lie on the same indifference surface since each is in  $\Omega$ . They all lie on the indifference surface corresponding to utility level  $E\{U(W + \tilde{Z})\}$ , call it  $\ell$ . Consider the vector  $Y$  defined as a point in the interior of the convex hull of the  $T$  vectors  $W - \Pi^t$ ,  $t=1, \dots, T$ :

$$(2.19) \quad Y = \sum_{t=1}^T \lambda_t (W - \Pi^t) \quad \text{where} \quad \sum_{t=1}^T \lambda_t = 1 \quad \text{and} \quad \lambda_t > 0 \quad \text{for all } t.$$

But then  $Y = W - \sum_{t=1}^T \lambda_t \Pi^t$ . The T-element vector  $\Pi^0 = \sum_{t=1}^T \lambda_t \Pi^t$  is an internal convex combination of T semipositive vectors each having the positive element in a different location. That is,  $\Pi^0$  lies in the interior of the convex hull of the T semipositive  $\Pi^t$ 's, no two of which have the positive element in the same location. Hence,  $\Pi^0 > 0$ , that is,  $\Pi^0$  is a positive vector.

Quasi-concavity of the utility function (diminishing marginal rates of substitution) was one of the assumptions made earlier. For any real k, the set of W-vectors such that  $U(W) \geq k$  is convex:

$$(2.20) \quad \theta(k) = \{W | U(W) \geq k\} \quad \text{is convex for any real } k .$$

But  $U(W - \Pi^t) = E\{U(W + \tilde{Z})\} = \ell$  for all t. Hence each  $W - \Pi^t$  is an element of  $\theta(\ell)$ . The definition of Y in (2.19) shows that it is an internal convex combination of the T  $W - \Pi^t$  vectors. Since  $\theta(\ell)$  is a convex set containing the T  $W - \Pi^t$  vectors, it therefore follows that  $Y \in \theta(\ell)$  and hence that  $U(Y) \geq E\{U(W + \tilde{Z})\}$ . One obtains, then,

$$(2.21) \quad U(W - \Pi^0) \geq E\{U(W + \tilde{Z})\} .$$

If the equality holds in (2.21) the proof is complete since a boundary insurance policy the decisionmaker would buy that has a positive premium in each period, namely,  $\Pi^0(W, \tilde{Z})$ , has been found.

Suppose, on the other hand, that the inequality holds in (2.21),  $U(W - \Pi^0) > E\{U(W + \tilde{Z})\}$ . An argument similar to that used in proving Theorem 2.1 is now employed. Due to the assumption of positive marginal utility for dollars in any period and the assumption that  $U(W)$  is at least twice continuously differentiable, by decreasing a single element of  $W - \Pi^0$  sufficiently one can obtain a vector  $W^*$  such that  $U(W^*) = E\{U(W + \tilde{Z})\}$ . But decreasing one element of  $W - \Pi^0$  while leaving all others unchanged can be accomplished by increasing one element of  $\Pi^0$  holding all others

constant. In short, one obtains a new vector  $\Pi^* \geq \Pi^0$ , with equality the case for all but one element, and

$$(2.22) \quad U(W - \Pi^*) = E\{U(W + \tilde{Z})\} .$$

Since  $\Pi^* \geq \Pi^0 > 0$ , a positive "boundary" risk-premium vector, namely,  $\Pi^*$ , has been found.

Because of the assumed smoothness of the utility function, there will generally be an infinite number of "boundary" insurance policies with positive premiums to be paid in each period that the decisionmaker would be willing to buy. The proof of Corollary 2.2 shows that there will always exist at least one such policy.

To summarize, then, an individual is a risk averter if and only if his multi-period utility function is concave and if and only if there always exists a set of "all-in-one-period" insurance policies he would be willing to purchase -- one member of the set corresponding to each period of his horizon. The questions that now come to mind are: "Can one extend the concepts of decreasing and increasing risk aversion to the multiperiod case?" and "If so, how does one measure whether an individual is decreasingly or increasingly risk-averse in the multiperiod context?" In the next section, answers are provided to both of these questions.

## 2.5. Decreasing Risk Aversion in the Multiperiod Case

### 2.5.A. The Definition of Decreasing Risk Aversion

As the individual decisionmaker's asset vector changes one might well expect that the insurance policies he would be willing to purchase to avoid a given risk vector would also change. In particular, as in the single-period case, decreasing risk aversion seems a most plausible property for a utility function to possess if it is to be used in making decisions under risk. Intuitively, decreasing risk aversion has the same meaning in the multiperiod case as in the single-period model.

As the individual's assets increase so that he becomes better off (as a result of the assumed positive marginal utility of dollars in any period at every point in assets space), he will want to pay less for insurance against a given vector of risks.

Formalizing this intuitive definition is not, however, a straightforward matter. The stumbling block, again, is the multiplicity of insurance policies a decisionmaker would be willing to purchase in order to avoid a given set of risks when his consumption incomes in the T periods are at given levels. The task becomes especially difficult since one would like to define this property without explicitly introducing the utility function of the insurance company or the characteristics of the other decisionmakers in the economy. Clearly, if the insurance company's utility function and the utility functions, asset vectors, and risk vectors of all the other decisionmakers to whom the company sells policies were given, one could discuss the changes in the solution obtained in (2.7) for the particular individual's insurance policy as one or more elements of his asset vector and only his asset vector increased. One could then define decreasing risk aversion as the case in which no premium of the individual's policy increases and at least one decreases as his consumption-income vector is incremented by a semipositive T-element vector. Similarly, increasing risk aversion would characterize the utility function if no premium of the policy decreased and at least one increased as the asset vector was increased in this way. The analogous definition for constant risk aversion should be clear: each element of  ${}_i\Pi$  stays the same as  ${}_iW$  increases.

An interesting feature of the multiperiod case that emerges in these definitions -- and that will also characterize the definitions to be used in what follows -- is the fact that decreasing, constant, and increasing risk aversion are not mutually exhaustive possibilities under these definitions. In the case of a single-stage utility function, they were mutually exclusive and exhaustive

possibilities. In the multiperiod case, due to the multidimensional nature of the insurance policy, while increasing, constant, and decreasing risk aversion remain mutually exclusive, they do not exhaust the entire set of possibilities. For example, if several  $\pi_t$  elements increased and several decreased as  $W$  increased, the individual would be neither increasingly risk-averse nor decreasingly risk-averse nor constantly risk-averse under the definitions just given. Quite clearly, though, he would not be exactly as averse to risk as he was before.

The fundamental problem with the definitions of decreasing and increasing risk aversion just given in terms of the solution for  $\Pi$  in (2.7) is their dependence upon the solution of that entire problem. One would hope to be able to define increasing and decreasing risk aversion as properties of the individual's behavior, independent of the particular economy in which he exists. The definitions in terms of how the solutions to (2.7) change as one set of parameters -- the particular individual's set of consumption incomes -- changes, however, clearly depend upon the insurance company's utility function, the other individuals' utility functions and their assets and risks.

On the other hand, it is also clear that one cannot simply employ the Hessian of the individual's utility function to answer questions about increasing and decreasing risk aversion. There is the obvious question of how to use the Hessian to detect increasing or decreasing risk aversion. But even if one could decide on a measure -- for example, some function of the principal minors of the Hessian -- the elements of the Hessian depend on the units in which utility is measured. As was the case with  $u''(W)$  in the single-period case, if  $U(W)$  were multiplied by a constant the individual's behavior would not be altered but his "degree of risk aversion", as measured by the function of the Hessian, would appear to be different.

Once again, in an attempt to resolve a problem in discussing attitudes toward risk in the multiperiod case, let us turn to the individual's

"all-in-one-period" insurance policies. Decreasing risk aversion in the multiperiod sense will be defined as follows.

Definition. A decisionmaker with utility function  $U(W)$  is decreasingly risk-averse if as  $W$  increases, that is, as at least one element of  $W$  increases and none decrease,  $\Pi^t(W, \tilde{Z})$  is nonincreasing for all  $t$  and strictly decreasing for at least one  $t$ , for all risks,  $\tilde{Z}$ . That is, as the individual's assets increase, the maximum amount he would be willing to pay for an insurance policy calling for a premium in one and only one period would decrease for at least one such policy and not increase for any of them. A similar definition applies for increasing risk aversion with  $\Pi^t(W, \tilde{Z})$  nondecreasing for all  $t$  and strictly increasing for at least one  $t$ , for all risks  $\tilde{Z}$ , as  $W$  increases. It is hoped that the reader will agree that these definitions are quite reasonable in attempting to discuss changing degrees of risk aversion without specifying the entire economy in which the individual makes his decisions.

Having defined decreasing risk aversion in a multiperiod context in terms of the "all-in-one-period" insurance policies the decisionmaker would willingly purchase, it is desirable to characterize utility functions that display decreasing risk aversion so defined. The remainder of this chapter is directed to providing a set of conditions on the utility function  $U(W)$  that are sufficient for decreasing risk aversion in the multiperiod sense. The conditions and the proof that they are sufficient are contained in Theorem 2.2 with which the discussion will conclude. No set of necessary and sufficient conditions will be provided since the author has not been able to uncover any such set.

#### 2.5.B. Some Preliminaries to A Set of Sufficient Conditions

In order to be able to discuss the sufficient conditions for decreasing risk aversion, the type of risks considered will have to be limited. The restriction is the result not of logical necessity but, if you will, of empirical necessity, namely, an inability to consider more general risk vectors. Specifically, attention

will be restricted to the class of independent risks. This means the risk vector  $\tilde{Z}$  can be written as

$$(2.23) \quad \tilde{Z} = \sum_{t=1}^T \tilde{Z}^t$$

where  $\tilde{Z}^t$  is a vector each of whose elements is zero except the tth which is the random variable  $\tilde{z}_t$ .

The risk faced in a specific period is assumed to be independent of the risk faced in any other period. The distribution of  $\tilde{z}_t$  is unaffected by the realization of any other risk  $\tilde{z}_s$  for  $s \neq t$ . For example, if each of the risks was normally distributed with mean  $\mu_t$  and variance  $\sigma_{z_t}^2$ , the vector  $\tilde{Z}$  (the T-dimensional random variable  $\tilde{Z}$ ) would have the T-variate normal distribution with mean

$$\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_t \\ \vdots \\ \mu_T \end{bmatrix} \text{ and a diagonal variance-covariance matrix with the } \underline{tth} \text{ diagonal element}$$

being  $\sigma_{z_t}^2$ . The restriction of the class of risk vectors considered to such independent risks has been found necessary in order to derive meaningful results concerning multiperiod decreasing risk aversion.

The risk vector  $\tilde{Z}^t$  removes all of the risks the individual was originally supposed to face -- as given by  $\tilde{Z}$  -- except for the one occurring in the tth period. Each  $\tilde{Z}^t$  is a T-element vector with  $\tilde{z}_i^t = 0$  for all  $i$  except  $i=t$  and  $\tilde{z}_t^t = \tilde{z}_t$  where  $\tilde{z}_t$  is the tth element of the original risk vector  $\tilde{Z}$ . In the presentation of sufficient conditions for decreasing risk aversion and in the proof of their sufficiency, the occasion arises to introduce a set of asset vectors,  $W^t$  for  $t=1, \dots, T$ , that are derived from  $W$  in the same way the  $\tilde{Z}^t$  vectors are derived from  $\tilde{Z}$ . Each new asset vector  $W^t$  leaves the individual with the same amount of consumption income he originally had in the tth period but with nothing in any other period. Hence,



each  $W^t$  is a T-element vector with  $W_i^t = 0$  for all  $i \neq t$  and  $W_t^t = W_t$  where  $W_t$  is the tth element of the original asset vector  $W$ .

Before introducing a new set of random variables that are crucial to the statement and proof of Theorem 2.2, it is necessary to come to grips with and to resolve a notational problem. Since  $\Pi^t(W, \tilde{Z}) \in \Omega(W, \tilde{Z})$  by definition,

$$(2.24) \quad U(W - \Pi^t(W, \tilde{Z})) = E\{U(W + \tilde{Z})\} \quad \text{for all } t .$$

The assumptions of (a) positive marginal utility for each period's dollars, and (b) at least continuous second partial derivatives imply that for  $W$  given,  $U(W - \Pi^t(W, \tilde{Z}))$  is a strictly decreasing continuous function of  $\pi_t^t$  (the only nonzero element of  $\Pi^t$ ). Hence,  $\pi_t^t$  and thus  $\Pi^t$  are uniquely defined by (2.24). There exists a function -- not a correspondence -- associating a single  $\Pi^t(W, \tilde{Z})$ , for each  $t$ , with each  $(W, \tilde{Z})$  pair. Solving (2.24) for  $\Pi^t(W, \tilde{Z})$  one obtains

$$(2.25) \quad \Pi^t(W, \tilde{Z}) = W - U^{-1}[E\{U(W + \tilde{Z})\}] .$$

Our notation fails us here because  $U^{-1}[E\{U(W + \tilde{Z})\}]$  remains a correspondence while, in fact, the second term (the second vector) on the right-hand side of (2.25) can only assume one value if (2.25) is to be an equation. Specifically, it must be identical with  $W$  in all but the tth element, and that element must differ from the tth one of  $W$  by the amount  $\pi_t^t$ . To overcome this problem, introduce the following functional notation:

$$(2.26) \quad U^{-1}[p, \bar{W}, t] \text{ is defined as the T-element vector } W \text{ such that}$$

(i)  $U(W) = p$  and (ii)  $W_i$  equals  $\bar{W}_i$  for all  $i$  except  $i = t$  .

Then  $\Pi^t(W, \tilde{Z})$  is properly defined as

$$(2.27) \quad \Pi^t(W, \tilde{Z}) = W - U^{-1}[E\{U(W + \tilde{Z})\}, W, t] .$$

It is important to note that with  $W$  and  $t$  given, there is a one-to-one mapping between the values  $E\{U(W+\tilde{Z})\}$  and the vectors  $U^{-1}[E\{U(W+\tilde{Z})\}, W, t]$ . Hence, the notation in (2.27) informs us, as it should, that  $\Pi^t(W, \tilde{Z})$  is a function not a correspondence: with each  $(W, \tilde{Z})$  pair there is associated one and only one risk-premium vector that has zeros everywhere but in the  $t$ th element.

As a last preliminary before proceeding to the proof of Theorem 2.2, a new set of random variables is introduced. Let  $\tilde{h} = U(W+\tilde{Z})$ . The risk vector  $\tilde{Z}$  generates a vector of additions to the original asset vector with a certain probability distribution. This results in a new random variable  $W+\tilde{Z}$  that in turn yields a random variable describing the decisionmaker's utility. It is this stochastic variable describing the individual's (stochastic) utility level resulting from the risk vector  $\tilde{Z}$  that is denoted by  $\tilde{h}$ . The risk vector  $\tilde{Z}$  maps into a unique distribution for  $\tilde{h}$  but the converse is not true:  $U^{-1}(\tilde{h})$  is a correspondence not a function.

Similarly, define  $\tilde{h}_t = U(W^t + \tilde{Z}^t)$ . This says the following. Suppose that the decisionmaker has assets in the  $t$ th period and only in the  $t$ th period. Suppose, moreover, that the amount of assets in that period is the same as what he was allotted by the original consumption-income vector  $W$ . The decisionmaker is then confronted with a risk in the  $t$ th and only the  $t$ th period and that risk is identical with the  $t$ th risk he faced in the original risk vector,  $\tilde{Z}$ . The vector  $\tilde{Z}^t$  generates an addition to his  $t$ th period assets and thus yields the new stochastic variable  $W^t + \tilde{Z}^t$ . This new stochastic variable then generates a distribution of the individual's utility level. It is the random variable with this distribution of the utility level that is called  $\tilde{h}_t$ .

With  $W$ , and hence  $W^t$  given, the risk vector  $\tilde{Z}^t$  maps into a unique distribution for  $\tilde{h}_t$ . In addition, since each period's assets are assumed to have positive marginal utility, with  $W^t$  given each  $\tilde{h}_t$  maps into a unique risk vector  $\tilde{Z}^t$  since  $\tilde{Z}^t$  contains only one possibly nonzero element. Hence,

$$(2.28) \quad \tilde{h}_t = U(W^t + \tilde{Z}^t) \quad \text{and} \quad W^t + \tilde{Z}^t = U^{-1}[\tilde{h}_t, W^t, t] .$$

Similarly, with  $\tilde{Z}^t$  given,  $W^t$  maps into a unique distribution for  $\tilde{h}_t$  and by the assumed positive marginal utility of each period's assets, each distribution of  $\tilde{h}_t$  maps into a single  $W^t$ . Finally, if attention is restricted -- as it is here -- to actuarially neutral risks,  $E(\tilde{Z}) = 0$ , then there is a one-to-one mapping between distributions of  $\tilde{h}_t$  and  $(W^t, \tilde{Z}^t)$  pairs.

### 2.5.C. Risk Balancing Over Time

With these new definitions of variables and functions, one can proceed to the statement of a set of sufficient conditions for decreasing risk aversion in the face of independent risks. First, we prove the following lemma upon which Theorem 2.2 relies and which helps in understanding one of the conditions imposed on  $U(W)$  in the theorem.

LEMMA 2.3. The function  $U(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t])$  is a convex function of the  $\tilde{h}_t$  variables if and only if

$$(2.29) \quad U(W - \Pi^t(W, \tilde{Z})) \geq U(W - \sum_{i=1}^T \Pi^i(W^i, \tilde{Z}^i)) \quad \text{for all } t.$$

PROOF: (i)  $U(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t])$  convex in the  $\tilde{h}_t$  variables  $\rightarrow$  (2.29)

If  $U(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t])$  is convex in the  $\tilde{h}_t$  (random) variables, it follows from

(2.15) in Lemma 2.1, concerning concave and convex functions, that

$$(2.30) \quad E\{U(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t])\} \geq U(\sum_{t=1}^T U^{-1}[E(\tilde{h}_t), W^t, t]) .$$

Since, from (2.28),  $W^t + \tilde{Z}^t = U^{-1}[\tilde{h}_t, W^t, t]$  and by definition  $\sum_{t=1}^T W^t = W$  and

$\sum_{t=1}^T \tilde{Z}^t = \tilde{Z}$ , it follows that

$$(2.31) \quad E\{U(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t])\} = E\{U(W+\tilde{Z})\} .$$

With  $\tilde{h}_t = U(W^t + \tilde{Z}^t)$ , it follows from the definition in (2.27) that  $W^t - \Pi^t(W^t, \tilde{Z}^t) = U^{-1}[E(\tilde{h}_t), W^t, t]$ . Hence, noting again that  $W = \sum_{t=1}^T W^t$ , one obtains

$$(2.32) \quad U(\sum_{i=1}^T U^{-1}[E(\tilde{h}_i), W^i, i]) = U(W - \sum_{i=1}^T \Pi^i(W^i, \tilde{Z}^i)) .$$

(The index was changed from  $t$  to  $i$  in going from the right-hand side of (2.30) to (2.32) for a reason of convenience that will become clear presently.) Substituting the results in (2.31) and (2.32) into (2.30), one finds

$$(2.33) \quad E\{U(W+\tilde{Z})\} \geq U(W - \sum_{i=1}^T \Pi^i(W^i, \tilde{Z}^i)) .$$

The statement in (2.29) then follows immediately from (2.33) because, from the definition of the individual's "all-in-one-period" policies, as made explicit in (2.24),

$$(2.24) \quad U(W - \Pi^t(W, \tilde{Z})) = E\{U(W+\tilde{Z})\} \quad \text{for every } t .$$

The proof of part (i) is thus complete.

$$(ii) \quad (2.29) \rightarrow U(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t]) \text{ is convex in the } \tilde{h}_t \text{ variables.}$$

Note first that if

$$(2.34) \quad U(W - \Pi^t(W, \tilde{Z})) \geq U(W - \sum_{i=1}^T \Pi^i(W^i, \tilde{Z}^i)) ,$$

then it is true for all  $t$  since, by definition, each  $U(W - \Pi^t(W, \tilde{Z})) = E\{U(W+\tilde{Z})\}$  and hence they are equal to each other. Given (2.29), then, the statement in (2.33) follows immediately. All the steps used in the proof of part (i) are reversible and one is led back to (2.30). With the initial asset vector  $W$  given,

$U(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t])$  is a function of the  $\tilde{h}_t$  variables alone and the result of Lemma 2.1 shows that (2.30) is equivalent to the convexity of  $U(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t])$  in the  $\tilde{h}_t$  variables. The proof of part (ii) is complete.

The inequality in (2.29) has an interesting and straightforward interpretation, which, as a result of Lemma 2.3, helps in understanding the convexity condition imposed upon  $U(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t])$  in Theorem 2.2. It expresses the decisionmaker's preference for an insurance policy that allows him to consider simultaneously all the risks he faces in all periods and to decide on a premium or set of premiums to cover such risks rather than having to insure against each risk out of the assets of the particular period in which the risk must be confronted. A decisionmaker who shows such a preference will be called a "risk balancer over time", or in what follows simply a "risk balancer".

Assume there were two insurance companies from whom the decisionmaker could buy a policy. Suppose the first company, Allrisk Incorporated, offered the individual a policy in which he paid a certain premium in the  $t$ th period which insured him against his  $T$  risks. The second company, One-At-A-Time Incorporated, makes available a different type of insurance. It tells the decisionmaker that in insuring against a given period's risk, it does not want to consider his assets in any other period or the other risks he faces. Instead, One-At-A-Time wants each risk insured against separately and strictly out of the assets the decisionmaker possesses in the period in which the risk occurs. If the individual in question is a risk balancer he will purchase his policy from Allrisk while if he is not a risk balancer, One-At-A-Time will have gained itself a customer. This pairing of customers with companies follows because Allrisk is offering the individual the policy  $\Pi^t(W, \tilde{Z})$  while One-At-A-Time is offering him the policy  $\sum_{i=1}^T \Pi^i(W^i, \tilde{Z}^i)$ .

It should be noted that the risk-balancing property does not imply that the individual would necessarily prefer an insurance policy under which he would pay the whole premium in one period to one in which he would pay a premium in each period. On the contrary, Corollary 2.2 to Theorem 2.1 showed that for any strict risk averter there is at least one boundary insurance policy he would purchase that involves the payment of positive premiums in every period. Call this policy  $\Pi^*$ , as was done in the proof of the corollary. Then, since  $U(W - \Pi^*(W, \tilde{Z})) = E\{U(W + \tilde{Z})\}$  and  $U(W - \Pi^t(W, \tilde{Z})) = E\{U(W + \tilde{Z})\}$ , one finds

$$(2.35) \quad U(W - \Pi^*(W, \tilde{Z})) \geq U(W - \sum_{i=1}^T \Pi^i(W^i, \tilde{Z}^i)) .$$

For a risk balancer, the policy  $\Pi^*(W, \tilde{Z})$  is also preferred to the policy offered by the One-At-A-Time company. The risk-balancing aspect of the decisionmaker's behavior does not mean he does not like to spread his premium payments over time. Rather, it means that he does not want to have to meet each risk from the assets of the particular period in which the risk occurs.

This risk-balancing property would seem to be a natural property of rational behavior under risk in a multiperiod setting. An alternative definition of an intertemporal risk balancer might be noted, although it will not be used in what follows.<sup>19</sup> This "tightened" definition characterizes a risk balancer as a decisionmaker for whom

$$(2.36) \quad U(W - \Pi^t(W, \tilde{Z})) \geq U(W - \sum_{i=1}^T \Pi^i(W, \tilde{Z}^i)) \quad \text{for all } t .$$

Once again, he prefers the insurance policy offered by Allrisk Incorporated. The policy offered by the One-At-A-Time company is, however, different than it was before. Now One-At-A-Time tells the individual that in insuring a given period's risk it is

---

<sup>19</sup>I am indebted to Professor T. C. Koopmans for having suggested this alternative in a private letter.

willing to consider the fact that he has assets in other periods, but it still wants each period's risk insured against separately. It can be shown, by a proof similar to that of Lemma 2.3 that this alternative definition of a risk balancer is equivalent to the following condition:

$$(2.37) \quad U((1-T)W + \sum_{i=1}^T U^{-1}[U(W+Z^i), W, i]) \quad \text{is a convex function of the} \\ U(W+Z^i) \text{ variables, } i=1, \dots, T .$$

2.5.D. Sufficient Conditions for Decreasing Risk Aversion  
in the Multiperiod Case

The preliminaries completed, let us now turn to the presentation of a set of conditions sufficient for decreasing risk aversion in the face of independent risks and to a proof of their sufficiency. This is accomplished by Theorem 2.2.

THEOREM 2.2. If  $U(W)$  is the utility function of a strictly risk-averting decisionmaker, then the decisionmaker is decreasingly risk-averse in the multiperiod sense with respect to independent risks if the following conditions are satisfied.

$$(2.38) \quad U\left(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t]\right) \text{ is convex in the } \tilde{h}_t \text{ variables;}$$

$$(2.39) \quad dU\left(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t]\right) \text{ is convex in the } \tilde{h}_t \text{ variables}$$

where the differential is with respect to nonnegative increments in the  $T$  initial consumption incomes;

$$(2.40) \quad \text{Setting all } W_t = 0 \text{ except } W_i \text{ in } U(W), \quad \frac{\partial^3 U}{\partial W_i^3} \cdot \frac{\partial U}{\partial W_i} \geq \left(\frac{\partial^2 U}{\partial W_i^2}\right)^2$$

and this is true for all  $i$ .

That is, if  $U(W)$  satisfies conditions (2.38), (2.39), and (2.40), then as  $W$  increases in at least one element and decreases in none,  $\Pi^t(W, \tilde{Z})$  is nonincreasing for all  $t$  and strictly decreasing for at least one  $t$  for all independent risks  $\tilde{Z}$ .

PROOF: Suppose the initial consumption-income vector is  $\bar{W} = (\bar{w}_1, \bar{w}_2, \dots, \bar{w}_T)$ . From the definition of the  $t$ th all-in-one-period insurance-premium vector, at the initial position

$$(2.41) \quad U(\bar{W} - \Pi^t(\bar{W}, \tilde{Z})) = E\{U(\bar{W} + \tilde{Z})\} \quad \text{for all } t.$$

Suppose the consumption-income vector increases by a set of incremental changes  $d\bar{w}_i \geq 0$  for all  $i$  and  $d\bar{w}_i > 0$  for at least one  $i$ . Since the condition in (2.41) must hold at both the initial and final positions, that is, at  $\bar{W}$  and at  $\bar{W} + d\bar{W}$ , for small enough  $d\bar{W}$  one obtains

$$(2.42) \quad d[U(\bar{W} - \Pi^t(\bar{W}, \tilde{Z}))] = d[E\{U(\bar{W} + \tilde{Z})\}] \quad \text{for each } t,$$

where the differential is with respect to the  $\bar{w}_i$ 's.

Since the differential and expectations operators are commutative, the equation in (2.42) may be rewritten as

$$(2.43) \quad d[U(\bar{W} - \Pi^t(\bar{W}, \tilde{Z}))] = E\{dU(\bar{W} + \tilde{Z})\} \quad \text{for each } t.$$

But

$$(2.44) \quad d[U(\bar{W} - \Pi^t(\bar{W}, \tilde{Z}))] = \sum_{i=1}^T \left( \frac{\partial U}{\partial \bar{w}_i} \right)_{\bar{W} - \Pi^t} d\bar{w}_i - \sum_{i=1}^T \left( \frac{\partial U}{\partial \bar{w}_t} \right)_{\bar{W} - \Pi^t} \frac{\partial \pi_t^t}{\partial \bar{w}_i} d\bar{w}_i,$$

where a subscript on a partial derivative (or on a differential in what follows) indicates the point at which it is to be evaluated. Notationally, we write  $\Pi^t(\bar{W}, \tilde{Z})$ , the only possibly nonzero element of which is  $\pi_t^t$ , as  $\Pi^t$  when no confusion is possible.



With regard to the last sum on the right-hand side of (2.44), note that  $(\frac{\partial U}{\partial \bar{W}_t})_{\bar{W}-\Pi^t}$  is clearly independent of  $i$ , and that since only the  $\bar{W}_i$  elements are changing,

$$d\pi_t^t = \sum_{i=1}^T \frac{\partial \pi_t^t}{\partial \bar{W}_i} d\bar{W}_i. \quad \text{Hence, (2.44) is equivalent to}$$

$$(2.45) \quad d[U(\bar{W}-\Pi^t(\bar{W}, \tilde{Z}))] = [dU(\bar{W})]_{\bar{W}-\Pi^t} - \left(\frac{\partial U}{\partial \bar{W}_t}\right)_{\bar{W}-\Pi^t} d\pi_t^t \quad \text{for each } t,$$

$$\text{since } \sum_{i=1}^T \left(\frac{\partial U}{\partial \bar{W}_i}\right)_{\bar{W}-\Pi^t} d\bar{W}_i = [dU(\bar{W})]_{\bar{W}-\Pi^t}.$$

From (2.28), it follows that  $\bar{W}+\tilde{Z} = \sum_{i=1}^T (\bar{W}^i+\tilde{Z}^i) = \sum_{i=1}^T U^{-1}[\tilde{h}_i, \bar{W}^i, i]$ . Using

this fact and the result in (2.45), the equation in (2.43) can be rewritten as

$$(2.46) \quad \left(\frac{\partial U}{\partial \bar{W}_t}\right)_{\bar{W}-\Pi^t} d\pi_t^t = [dU(\bar{W})]_{\bar{W}-\Pi^t} - E\{d[U(\sum_{i=1}^T U^{-1}[\tilde{h}_i, \bar{W}^i, i])]\}.$$

The assumption in (2.39) combined with the earlier result (2.15) on the expected value of convex functions imply that

$$(2.47) \quad E\{d[U(\sum_{i=1}^T U^{-1}[\tilde{h}_i, \bar{W}^i, i])]\} \geq dU(\sum_{i=1}^T U^{-1}[E(\tilde{h}_i), \bar{W}^i, i]).$$

Hence,

$$(2.48) \quad \left(\frac{\partial U}{\partial \bar{W}_t}\right)_{\bar{W}-\Pi^t} d\pi_t^t \leq [dU(\bar{W})]_{\bar{W}-\Pi^t} - dU(\sum_{i=1}^T U^{-1}[E(\tilde{h}_i), \bar{W}^i, i]) \quad \text{for each } t.$$

From the definition of insurance-premium vectors, it is known that

$$U(\bar{W}^i - \Pi^i(\bar{W}^i, \tilde{Z}^i)) = E\{U(\bar{W}^i + \tilde{Z}^i)\} = E(\tilde{h}_i) \quad \text{and thus}$$

$$(2.49) \quad \bar{W}^i - \Pi^i(\bar{W}^i, \tilde{Z}^i) = U^{-1}[E(\tilde{h}_i), \bar{W}^i, i] \quad \text{for each } i.$$

Denoting  $\Pi^i(\bar{W}^i, \tilde{Z}^i)$  by  $\Pi^{*i}$  and, for future reference, its  $i$ th element by  $\pi_i^{*i}$ ,

inequality (2.48) becomes

$$(2.50) \quad \left( \frac{\partial U}{\partial \bar{W}_t} \right)_{\bar{W}-\Pi^t} d\pi_t^t \leq [dU(W)]_{\bar{W}-\Pi^t} - dU \left( \sum_{i=1}^T [\bar{W}^i - \Pi^{*i}] \right) \text{ or}$$

$$(2.51) \quad \left( \frac{\partial U}{\partial \bar{W}_t} \right)_{\bar{W}-\Pi^t} d\pi_t^t \leq [dU(W)]_{\bar{W}-\Pi^t} - dU(\bar{W} - \sum_{i=1}^T \Pi^{*i}) \quad \text{for each } t,$$

since  $\bar{W} = \sum_{i=1}^T \bar{W}^i$  by the definition of  $\bar{W}^i$ .

But

$$(2.52) \quad dU(\bar{W} - \sum_{i=1}^T \Pi^{*i}) = \sum_{i=1}^T \left( \frac{\partial U}{\partial \bar{W}_i} \right)_{\bar{W}-\sum_{i=1}^T \Pi^{*i}} \left( 1 - \frac{\partial \pi^{*i}}{\partial \bar{W}_i} \right) d\bar{W}_i.$$

Note that since  $\Pi^{*i}$  is a function of  $\bar{W}^i$  and  $\tilde{Z}^i$  only, and since only  $\bar{W}^i$  (not  $\tilde{Z}^i$ ) is changing, the only partial derivative that had to be taken in the case of each  $\Pi^{*i}$  was the one of  $\pi^{*i}$  (the only nonzero element of  $\Pi^{*i}$ ) with respect to  $\bar{W}_i$  (the only nonzero element of  $\bar{W}^i$ ). Moreover, in this case,  $d\pi^{*i} = \left( \frac{\partial \pi^{*i}}{\partial \bar{W}_i} \right) d\bar{W}_i$ . The expression for  $dU(\bar{W} - \sum_{i=1}^T \Pi^{*i})$  may therefore be written as

$$(2.53) \quad dU(\bar{W} - \sum_{i=1}^T \Pi^{*i}) = [dU(W)]_{\bar{W}-\sum_{i=1}^T \Pi^{*i}} - \sum_{i=1}^T \left( \frac{\partial U}{\partial \bar{W}_i} \right)_{\bar{W}-\sum_{i=1}^T \Pi^{*i}} d\pi^{*i},$$

since  $\sum_{i=1}^T \left( \frac{\partial U}{\partial \bar{W}_i} \right)_{\bar{W}-\sum_{i=1}^T \Pi^{*i}} d\bar{W}_i = [dU(W)]_{\bar{W}-\sum_{i=1}^T \Pi^{*i}}$ .

Using the result in (2.53), the inequality in (2.51) is equivalent to

$$(2.54) \quad \left( \frac{\partial U}{\partial \bar{W}_t} \right)_{\bar{W}-\Pi^t} d\pi_t^t \leq [dU(W)]_{\bar{W}-\Pi^t} - [dU(W)]_{\bar{W}-\sum_{i=1}^T \Pi^{*i}} + \sum_{i=1}^T \left( \frac{\partial U}{\partial \bar{W}_i} \right)_{\bar{W}-\sum_{i=1}^T \Pi^{*i}} d\pi^{*i} \quad \text{for each } t.$$

Lemma 2.3 has proved, however, that condition (2.38) is equivalent to

$$(2.55) \quad U(\bar{W} - \Pi^t(\bar{W}, \tilde{Z})) \geq U(\bar{W} - \sum_{i=1}^T \Pi^i(\bar{W}^i, \tilde{Z}^i)) \quad \text{for every } t.$$

Therefore, from the concavity of  $U(W)$  -- since the decisionmaker is assumed to be a strict risk averter -- and condition (2.38) it follows that

$$(2.56) \quad [dU(W)]_{\bar{W} - \Pi^t} < [dU(W)]_{\bar{W} - \sum_{i=1}^T \Pi^i(\bar{W}^i, \tilde{Z}^i)} \quad \text{for each } t.$$

Hence, combining (2.54) and (2.56) one has

$$(2.57) \quad \left( \frac{\partial U}{\partial W_t} \right)_{\bar{W} - \Pi^t} d\pi_t^t < \sum_{i=1}^T \left( \frac{\partial U}{\partial W_i} \right)_{\bar{W} - \sum_{i=1}^T \Pi^i} d\pi_i^{*i} \quad \text{for each } t.$$

The assumption in (2.40), however, ensures that  $\frac{\partial \pi_i^{*i}}{\partial \bar{W}_i} < 0$  for each  $i$ . This implication of (2.40) is best seen by considering a single-period utility function  $u_i(W_i)$  for the  $i$ th period. Pratt's work, summarized earlier, demonstrates that  $u_i(W_i)$  is decreasingly risk-averse if and only if  $u_i'(W_i)u_i'''(W_i) \geq [u_i''(W_i)]^2$ . But if all  $W_t$ 's except the  $i$ th are set equal to zero in  $U(W)$ , the result is a single-period utility function in  $W_i$  alone. The risk premium  $\pi_i^{*i} = \pi_i^{*i}(W_i, \tilde{Z}^i)$  is a decreasing function of  $W_i$  for all  $\tilde{Z}^i$  if and only if that resultant utility function meets the condition on the derivatives of the single-period utility function that was just given. Assumption (2.40) explicitly states that this condition is met. Hence, for each  $i$ ,  $\frac{\partial \pi_i^{*i}}{\partial \bar{W}_i} < 0$ . With  $d\bar{W}_i \geq 0$  for all  $i$  and  $d\bar{W}_i > 0$  for at least one  $i$ , then,  $d\pi_i^{*i} \leq 0$  for all  $i$  and  $d\pi_i^{*i} < 0$  for at least one  $i$  in (2.57).

Since consumption incomes in every period are assumed to have positive marginal utility, no matter what the asset vector, the right-hand side of (2.57) is thus negative. Finally, since  $\left( \frac{\partial U}{\partial W_t} \right)_{\bar{W} - \Pi^t} > 0$  by the assumption of positive marginal utility, it follows that

$$(2.58) \quad d\pi_t^t < 0 \quad \text{for each } t .$$

Since  $\pi_t^t$  is the only element of  $\Pi^t(\bar{W}, \tilde{Z})$  that is ever nonzero, it follows that for any set of differential increments in  $\bar{W}$ ,  $\Pi^t(\bar{W}, \tilde{Z})$  decreases for independent risks if  $U(W)$  satisfies conditions (2.38)-(2.40). This is true for every initial consumption-income vector,  $\bar{W}$ . That is, it is true of every point in the T-dimensional assets space. Hence, for any increase in  $\bar{W}$ , it is true that if  $U(W)$  satisfies conditions (2.38)-(2.40), each  $\Pi^t(\bar{W}, \tilde{Z})$  will decrease if  $\tilde{Z}$  is a vector of independent risks.

Note that the conditions given in (2.38)-(2.40) actually ensure something stronger than the definition of decreasing risk aversion required. The result in (2.58) shows that each  $\Pi^t(\bar{W}, \tilde{Z})$  will actually decrease while the definition, on the other hand, required only that no  $\Pi^t(\bar{W}, \tilde{Z})$  increase and that at least one decrease.

Before concluding this chapter, it is interesting to note the relationship between the set of sufficient conditions given in (2.38)-(2.40) for decreasing risk aversion in the multiperiod sense and Pratt's result for the single-period case. As Pratt shows,<sup>20</sup> a single-stage utility function  $u(W)$  is (strictly) decreasingly risk-averse if and only if

$$(2.59) \quad u'(u^{-1}(\tilde{q})) \text{ is a (strictly) convex function of } \tilde{q} \text{ or,} \\ \text{equivalently, } u'(W)u'''(W) \geq [u''(W)]^2 ,$$

where  $\tilde{q}$  is a random variable indicating a stochastic level of utility. If the horizon  $T$  is set equal to unity, the conditions in (2.38)-(2.40) reduce to Pratt's single-period condition given in (2.59).

---

<sup>20</sup>Pratt (1964, pp. 130-131).

First, there is no single-period analogue to condition (2.38) since, with only one period, there is no meaning to risk balancing over time. Nevertheless, formally setting  $T = 1$  in (2.38), the condition becomes  $U(U^{-1}[\tilde{h}_1, W_1^1, 1])$ , with  $W_1^1$  a single-element vector, is convex in  $\tilde{h}_1$ . But  $U(U^{-1}[\tilde{h}_1, W_1^1, 1]) = \tilde{h}_1$  as a result of the one-to-one nature of the mapping defined in (2.28). Since  $\tilde{h}_1$  is convex in itself, because it is linear in itself, the first condition of the theorem is met by any single-period utility function.

With  $T = 1$ , condition (2.39) requires that  $dU(U^{-1}[\tilde{h}_1, W_1^1, 1])$  be convex in  $\tilde{h}_1$  where the differential is with respect to positive increments in the only initial consumption income  $W_1$ . Since the mapping in (2.28) is one-to-one and since the vector involved does contain only one element, this is equivalent to requiring that  $du(u^{-1}(\tilde{h}_1))$  be convex in  $\tilde{h}_1$  for a positive differential change in  $W_1$ . But this means  $u'(u^{-1}(\tilde{h}_1))dW_1$  convex in  $\tilde{h}_1$  for  $dW_1 > 0$  or, therefore,  $u'(u^{-1}(\tilde{h}_1))$  convex in  $\tilde{h}_1$ . This is precisely the condition in (2.59).

Finally, with  $T = 1$ , the condition in (2.40) is exactly the second equivalent form of Pratt's single-period condition in (2.59).

Hence, if there is, in fact, only one period, Pratt's condition (2.59) and the sufficient conditions in (2.38)-(2.40) are identical.

## 2.6. Concluding Remarks

Before discussing the question of risk aversion over time, a brief summary of the results on attitudes toward risk in a one-period setting was given. An attempt was made to strengthen the case for the plausibility of decreasing risk aversion as a characteristic of rational behavior in the presence of risk. Certain phenomena that seemed to contradict the plausibility of decreasing risk aversion were reconciled with the rationality of this characteristic.

The discussion then turned to the issue of risk aversion and the multi-period utility function. The problems encountered when one moves from a

one-period horizon to a multiperiod one were discussed. In particular, it was seen that the source of the difficulties was the nonuniqueness of the risk-premium vector  $\Pi$ . Explicit introduction of an insurance company (a passive agent in the one-period case) with its own multiperiod utility function was then shown to be helpful in resolving the unanswered questions of the multiperiod case.

The possibility of learning something about the presence or absence of risk aversion and its varying degree from one asset level to another for the same decisionmaker without explicitly introducing an insurance company was then explored. The relationship between the multiperiod utility function's shape and the presence of risk aversion was shown to be the same as in the single-period case. The central role of the decisionmaker's "all-in-one-period" insurance policies was then demonstrated and several implications of risk aversion for such policies were proved. Making extensive use of these policies, a set of sufficient conditions was presented for a single decisionmaker to be decreasingly risk-averse in the face of a set of independent risks. The new concept of risk balancing over time was introduced as an important member of this set of sufficient conditions.

The results of this chapter will be very important in discussing the problem of capital budgeting under risk. In investigating this particular example of multiperiod decisionmaking in an environment of risk, the results concerning concavity of the utility function and the set of sufficient conditions for decreasing risk aversion will be extremely important. In particular, we will be concerned with the type of multiperiod utility function that is required in formulating a capital-budgeting problem if risk aversion and decreasing risk aversion characterize the decisionmaker. The form the final objective function takes when the expected-utility hypothesis is applied to such a utility function will have interesting implications about the decisionmaker's preferences among the several properties (mean, variance, skewness) of the relevant probability distributions.

But this is getting much too far ahead of the discussion. The next chapter presents an introductory view of the capital-budgeting problem with which this study is concerned, including a brief indication of the role that risk plays in the investment decision. After discussing the capital-budgeting problem under certainty in greater detail, we shall return to explore this decisionmaking problem in its risk environment. At that time, we will reap the benefits of the discussion in this chapter.

## CHAPTER 3

### THE CAPITAL-BUDGETING PROBLEM:

#### AN INTRODUCTORY VIEW

##### 3.1. Introduction

Having discussed the general problem of decisionmaking under risk in a multiperiod context, the present chapter turns to a specific case of such decisions: capital budgeting under risk. The chapter begins by defining, more precisely than Chapter 1 did, the capital-budgeting problem with which this study is concerned. The nature of the imperfect capital market the firm is assumed to face is discussed, with some attention being paid to the flexibility of the model used. It is shown how additional assumptions could enable the model to depict several different types of imperfect capital market. The chapter goes on to demonstrate how the capital-budgeting problem takes the form of a mathematical-programming model. The programming model of capital budgeting is then compared with analogous decisionmaking situations. Finally, the chapter discusses the role of risk in the capital-budgeting problem, as a preview to the more detailed discussion to be presented in Chapter 6. The chapter closes with some brief comments on the objective of the entire capital-budgeting discussion presented in this study.

##### 3.2. The Capital-Budgeting Problem More Precisely Defined

The "capital-budgeting problem", as it will be defined and considered in what follows, is the problem of allocating fixed budget dollars in several time periods among competing investment proposals. From the set of available investment opportunities, extending over the enterprise's planning horizon, the optimal



subset of projects to be undertaken must be chosen given the fact that the dollars available for use in every period are limited. The total amount of money available for investment purposes in any future period and the composition of the financing of that amount are taken as given. We inquire neither into how the total amounts available are determined nor into how the firm should raise capital, as, for example, between debt and equity issues.

Funds may be limited because the unit in question is a corporate division or a government agency whose budget is set by an upper echelon, the corporate headquarters in the first case, executive and legislative branches of government in the second. Or, upper bounds on the amount of money available for investment purposes in any period may exist because the firm may want to finance its projects mainly or only from internally generated funds. In some instances, the firm may have no recourse to the capital market -- or, more generally, the firm may have no recourse to a perfect capital market.

Some writers have taken the position that the long-term ceiling on the capital outlays of a firm is, in fact, set by the volume of its internally generated funds.<sup>1</sup> In contrast, other students of capital budgeting have argued that "the share of its income a corporation retains is not beyond the control of its management, and among the things we want from a capital budgeting model is guidance on whether the share of a corporation's income that is retained for investment should be raised or lowered."<sup>2</sup> The present author would agree, in principle, with the second pair of writers cited. But the ceteris paribus assumption made earlier -- specifically, that this study will not develop methods for setting the upper bound on total investment in each future period -- leaves us aligned in practice with Dean and Terborgh.

---

<sup>1</sup>Dean (1951, pp. 53-55); Terborgh (1949, pp. 228-229).

<sup>2</sup>Gordon and Shapiro (1956, p.147).

In the context of a study with the limited goal of considering investment-project selection, this assumption that period-by-period internal limits on investment spending are given is a reasonable one. The capital-budgeting problem remains complex and serious, definitely worthy of consideration and study, even when the expected future cash throw-offs generated by the firm and attributable to its ongoing operations are taken as given. For one thing, the interdependence of planning in different periods that results from the presence of multiple budgets and all its attendant complexity remains very much with us. The assumption simply freezes one possible set of variables -- the budget levels -- and abstracts from their determination so that the remaining part of the decision process may be better analyzed and understood.

There are, in fact, situations where the capital-budgeting problem arises in almost precisely the form just stated. Marschak, for example, writes of the French nationalized industries that "Each of the nationalized industries is regularly faced with the choice of an investment program given a fixed quantity of capital funds."<sup>3</sup> In more theoretical terms, Turvey assures us that the problem as set out above with given expenditure ceilings will be with us for quite a while. Speaking of public investment in roads, he writes

In practice, therefore, the questions "which roads?" and "how much roads?" are frequently separated. Perhaps they are separated too frequently, but even so, the first case, involving suboptimization (i.e., making the best lower-level decisions within a given framework provided by previous higher-level decisions) will inevitably continue to arise as long as administrative talent is limited. Somewhere or other down the line there have to be either minimum amounts or minimum periods of time within which subordinates have the task of choice delegated to them. The Ministry of Power could never, even if it wanted to, centralize all decisions about fuel investment projects, and the Board of an insurance company will always have to allocate investment funds between its property managers and its investment managers for finite periods.<sup>4</sup>

---

<sup>3</sup>T. Marschak (1960, p. 134).

<sup>4</sup>Turvey (1963, p. 94).

The period-by-period budgets thus serve as control devices that limit the firm's expenditures in each period. In this sense, they stand in contrast to budgets used as planning devices. Budgets of the latter type indicate the factor requirements at specific future dates and act as sources of information and guidelines for decisions to be taken at those later times.<sup>5</sup> In the capital-budgeting framework, it is the final capital-investment program that serves as such a planning budget.

### 3.3. The Imperfect Capital Market and the Planning Horizon

While funds limitations are a crucial aspect of the capital budgeting problem, it seems empirically more reasonable to allow the decisionmaking unit somewhat greater flexibility in its expenditures than its supply of internally generated funds alone would permit. In what follows, the enterprise is freed from what one could call a "pure" capital rationing situation where it has for its own use and only for its own use fixed amounts of money in each period. The model developed in this study allows the enterprise to borrow and to lend but still forces it to operate in a world of imperfect capital markets. Now there is need not only to consider the interrelationships between physical investments alone, but also between financing decisions and decisions concerning physical projects.

Specifically, the firm has available the period-by-period anticipated future cash throw-offs its present assets will generate. In addition it can borrow funds and lend funds on the market, each at a constant rate of interest, but the borrowing rate is higher than the rate of interest it receives on money it lends. The difference between the two rates is the price that must be paid to offset the labor and transaction costs involved in bringing together the borrower and the lender. It is the return that must be paid to the intermediary for relieving the borrower and the lender of the costs and effort of locating one another and of bargaining once they

<sup>5</sup>This distinction between budgets as plans and budgets as controls is sharply drawn in Stedry (1960, pp. 1-5).

have found each other. Lastly, while it can lend as much as it wants to, within the constraint of its anticipated internally generated funds, it is subject to an absolute borrowing limit in each period. Associated with each budget period, there is a maximum amount of debt the enterprise can incur, and this upper bound may change exogenously from one period to the next.

Before proceeding any further, it ought to be made clear that no attempt is being made to cover the entire spectrum of capital-budgeting models. For some corporations or corporate divisions or government agencies, the pure-capital-rationing model may be appropriate. Other corporations or their divisions may find the model based on the imperfect capital markets just described -- the one to be studied in detail here -- more applicable. Yet other enterprises might operate in capital-market environments much more complex than either of the two models stated briefly above. In fact, more likely than not, most economic units planning capital-investment programs do make their decisions in environments that would be very difficult to capture in a tractable model of the type to be used here. Hence, rather than attempting the impossible -- or at least the nearly impossible -- a single model of the capital-budgeting problem has been chosen for use in this work. It is a model which seems to come closer to the real-world environment of capital-budgeting decisions than does the pure-capital-rationing model, and it also is flexible enough to enable it to be extended to more complex and perhaps more realistic situations.<sup>6</sup>

The period-by-period borrowing limits may be due to any one of a number of factors. They may represent leverage constraints that are either self-imposed or induced by the tests bankers and investment analysts apply in evaluating firms. In such instances, the borrowing limitations constrain the financial management to maintain certain ratios between the amount of debt outstanding and the equity value

---

<sup>6</sup>Weingartner (1963, Chapters 8 and 9) presents a series of capital-budgeting models for perfect and imperfect capital markets. The model with absolute borrowing limits is discussed in Section 9.1. The differences between Weingartner's models and the one developed here will be made explicit in a later chapter.

of the firm, since it is assumed in the present model that no new common stock is issued over the horizon period.<sup>7</sup> That is, with the equity allowed to increase only by the amount of retained earnings, a constraint on borrowing is equivalent to a leverage constraint.<sup>8</sup>

Another source of borrowing limits may be the existence of maximum amounts of trade credit that suppliers would willingly extend. Here the borrowing limit is imposed directly by the suppliers of funds. Borrowing limits may also have their origin external to the enterprise in the case of public utilities that were put under the jurisdiction of the Securities and Exchange Commission by the Public Utilities Holding Act of 1935. For these companies, although they are now few in number, the Securities and Exchange Commission has the power to prescribe capital ratios.<sup>9</sup> State regulatory bodies have assumed comparable powers with regard to utilities under their jurisdiction.

A brief look at the asserted flexibility of the model, which will be central to the ensuing discussion, seems worthwhile. First, the case of pure capital rationing is a specific subcase of the proposed model with divergent interest rates and borrowing limits. To obtain the model of pure capital rationing, one simply restricts the amount of debt each period to nonpositive quantities and sets the lending

---

<sup>7</sup>For a model that incorporates the amount and timing of new equity financing, see Weingartner (1963, Section 9.5).

<sup>8</sup>An alternative way of introducing a leverage constraint on the firm would be the following. Assume the firm has settled upon an optimal debt-equity ratio, either of its own volition or as a result of the bankers' and investment analysts' debt-equity tests. At the start of the investment program, its debt-equity ratio is assumed to be at this optimal level. Hence, when it raises additional funds over the course of the capital-investment program's horizon, it will obtain these marginal funds in the same optimal proportion. Now reinterpret "the borrowing rate" mentioned above as the average rate at which the corporation can obtain funds, where the average is determined by the proportion of its funding that is in the form of debt issues and the proportion that is in the form of new equity. It is the real rate at which funds are available. There is now no need for the borrowing limits as the leverage constraint is effected in the definition of the borrowing rate. The present study will, however, use the formulation described in the text.

<sup>9</sup>Weingartner (1963, p. 161).

rate to zero. The enterprise would not be able to borrow and it would have no desire to lend. Alternatively, to obtain the model of pure capital rationing, one could set the borrowing rate at a very high level such that borrowing would never be worthwhile and set the lending rate to zero so that there would be no desire to lend. Whichever path is chosen, the corporation would plan its capital-investment program without turning to the capital market.

The case diametrically opposed to pure capital rationing, namely, the case of a perfect capital market, is also easily derived from the model that will be discussed in this study. First, one sets the borrowing rate equal to the lending rate. Then one removes the borrowing limits, for example, by setting them to infinity or to some amount greater than the firm could conceivably want to borrow. The enterprise can then borrow or lend unlimited amounts at fixed and equal interest rates: it has access to a perfect capital market.

Moving in the direction of generalization rather than specialization, the model central to this study can be converted into a capital-market environment where the borrowing rate rises as the amount borrowed increases, that is, a market with an upward-sloping supply-of-funds schedule. Instead of a single borrowing activity in each period there are several borrowing activities, each with a different interest rate applicable. Moreover, there is not just one borrowing limit in each period, but rather there is an upper bound on the amount that can be borrowed at each of the different interest rates. Since it is in the firm's interest to obtain funds at the lowest possible cost, it will borrow whatever it needs at the lowest interest rate until it exhausts that supply, as given by the borrowing limit. If it requires even more money to pursue an optimal investment program, the corporation will then proceed to borrow at the next highest interest rate. It will continue in this fashion, exhausting the supply of funds available at a given rate before going on to borrow at the next level, until it has obtained the

requisite funds. In this market, then, for each period the marginal cost of borrowed funds is a step function -- a new step beginning each time a borrowing limit is exhausted. The result is a supply (average cost) schedule for borrowed funds that is piecewise linear and rises as the amount borrowed increases.<sup>10</sup>

Another form of generalization that could easily be applied to the divergent-rates, borrowing-limits model is that of intertemporal differences in the market interest rates. It has tacitly been assumed in describing the several possible models that the lending rate and the borrowing rate (or the schedule of borrowing rates in the case of a rising supply schedule) are the same from one period to the next. This assumption is easily modified, and the model thereby generalized, if time subscripts are appended to the interest rates. In what follows it shall be assumed, however, that the borrowing rate is the same for all periods and that the lending rate is also constant over time, in addition to each of them being constant within each period.

Within the setting of the imperfect capital market that has been described, the corporation seeks to solve its capital-budgeting problem. Taking account of its anticipated cash throw-offs and the capital-market options available to it, the firm chooses the optimal subset of investments to be made from among the available opportunities extending over the planning horizon. The length of the planning horizon depends, of course, on the individual decisionmaking unit. In the discussion to follow, the number of future periods considered in planning the investment program is taken as a given.

In choosing the horizon, one would like to follow certain guidelines. Specifically, the decisionmaker would like to set the horizon  $T$  at a point in time "such that the set of accepted projects having outlays or revenues in year  $T$  or sooner are exactly the same whether the model makes use of an infinite horizon or

---

<sup>10</sup> For a fuller discussion of such a model, see Weingartner (1963, Section 9.3).

a horizon set at T" or a point in time "such that the decisions which call for implementation before this date will be exactly the same, whether or not events past that moment are treated explicitly or implicitly...."<sup>11</sup> Unfortunately, as Weingartner goes on to tell us, "In dynamic models in general such a horizon does not necessarily exist, or there may be many of them. If there are several, the earliest having this property may be designated as the preferred one"<sup>12</sup> because of the problems entailed in collecting data about prospective investments. In the case of a multiplicity of potential horizon periods, Weingartner's choice seems sound. Clearly, the real difficulty arises when no period T satisfies the desiderata for a horizon cut-off point. The present study, however, has no advice to offer on this matter. Instead, it circumvents the problem by assuming that a horizon period T has been given by the decisionmaking unit, whether or not the horizon so chosen satisfies the conditions set out above.

#### 3.4. The Capital-Budgeting Problem As a Programming Model: A Verbal Presentation

The definition of the capital-budgeting problem given here has emphasized that the investment decision involves selecting a subset of projects to be pursued as opposed to accepting or rejecting individual proposals. With the exception of the growing programming literature on capital budgeting<sup>13</sup> and the well-known work of J. Hirshleifer, especially his article "On the Theory of Optimal Investment Decision,"<sup>14</sup> writers in the field of capital budgeting have concentrated on choosing the individual projects that should be included in the investment program. Despite the demonstrations by Hirshleifer and the programming-oriented writers of the grave shortcomings of the project-by-project approaches to investment decisionmaking, the

<sup>11</sup>Ibid., p. 153.

<sup>12</sup>Ibid., pp. 153-154.

<sup>13</sup>Works in the programming literature that have rejected this project-by-project approach to capital budgeting include: Baumol and Quandt (1965); Charnes, Cooper, and Miller (1959); Cord (1964); Hillier (1964); Lintner (1965); Manne (1966); Marglin (1963); Naslund (1964, 1966); Reiter (1963); and Weingartner (1963, 1966).

<sup>14</sup>Hirshleifer (1958).



literature, by and large, has gone about composing capital programs by piecemeal selection of investment projects: accepting or rejecting project A and then turning to consider project B and so on. While such approaches are valid so long as the corporation exists in a world of certainty, perfect capital markets, continuous and continuously differentiable transformation functions, and independent investment projects,<sup>15</sup> once one leaves such a world, these approaches are incorrect. Exactly why they are invalid when these conditions are not met will be discussed in detail in Section 4.3. It is important to note, however, that neither the "real world" for which the rules were intended nor the capital-budgeting model of this study meet these requirements. Instead of erring by examining projects one-by-one in the model used here, it is necessary to concentrate on selecting the optimal group of projects. This is a fundamental property of the capital-budgeting problem considered in this study.

As stated here, the capital-budgeting problem is clearly an optimization problem. The criterion function to be optimized has not yet been explicitly introduced because, as shall be seen, determining what that criterion function ought to be will constitute a major task in what follows. Nevertheless, the firm's goal in undertaking capital investments is clearly to improve the position of some individual or group of individuals. We leave until later the job of providing answers to the questions: whose position and how is it defined. There is, however, a criterion function the value of which depends on the projects undertaken and the aim in choosing the proposals to accept is to maximize (or minimize, as the case may be) this function.

The enterprise is, however, not able to accept whichever projects it wants to in performing this maximization or minimization. On the contrary, the optimization takes place within the framework of certain constraints. Among these are the

---

<sup>15</sup>Two investment proposals are independent when the feasibility of accepting either of them is unaffected by the acceptance or rejection of the other.

dollar budget ceilings and the period-by-period borrowing limits discussed earlier. Since investment projects are generally discrete indivisible possibilities, the enterprise must also choose projects subject to the restriction that it does not, for example, decide to buy three-fourths of a new machine or two-fifths of a building of given size. In short, there will be restraints on the values the project variables can assume.

Physical interrelationships among projects will introduce additional constraints subject to which the firm must make its choice. Some proposals may be mutually exclusive: accepting one of them means that it is not feasible to undertake another or a set of others. On the other hand, the acceptance of one project may be contingent upon the acceptance of another. For example, a firm may decide to construct a building that can be used for different purposes regardless of what else the firm does, but it will not purchase a new set of equipment without first constructing the plant to house the equipment. Such interrelationships increase the number of restrictions within which the firm must plan its capital investments.<sup>16</sup>

Other facts of business life impose further constraints on the enterprise. For example, manpower constraints may also limit the firm's ability to choose. Such "personnel budgets" may arise in a variety of contexts. Weingartner mentions the development of new weapons systems as one place where they frequently appear.<sup>17</sup> Conversations the author has had with officials of several companies have indicated that in some instances it is, in fact, the human-capital ceiling rather than the financial budget that is most restrictive.

Two other resource limitations that may arise and reduce further the firm's maneuverability in planning are material limitations and the need to present specific types of collateral in order to make certain kinds of loans. In the former case, the enterprise simply does not have an unlimited supply -- or at least does not

---

<sup>16</sup>The formulation of the constraints corresponding to such interdependences will be given in detail in Chapter 5.

<sup>17</sup>Weingartner (1963, pp. 125-126).

have enough -- of a particular material resource, for example, land, to be able to carry out whatever investment program it desires. The added type of financial constraint involving collateral restrictions can easily be incorporated into the model set out earlier. Such a constraint might occur if a loan were made, say, for the specific purpose of building a plant. While these funds could spill over and be used for other projects, certain minimal compliance in the direction of constructing a plant would be required if the firm was to obtain the money it desired.

In solving the capital-budgeting problem the firm thus seeks to maximize (or minimize) a yet unspecified objective function of a number of project variables subject to certain constraints on these project variables. The capital-budgeting problem is, consequently, a constrained optimization problem par excellence. The problem of investment-project selection is a mathematical-programming problem as it seeks an optimal allocation of limited resources. The remainder of this study will be concerned with the analysis of the capital-budgeting decision in its programming framework.

### 3.5. The Programming Model of Capital Budgeting and Its Analogues

Before going on to discuss the enterprise's knowledge of the environment within which it operates, consider briefly several other decisionmaking situations analogous to this programming formulation of the capital-budgeting problem. First, the capital-budgeting problem as presented here is very closely related to the standard consumer constrained-maximization problem. One major difference between the two is the discrete nature of the capital investments the firm makes as opposed to the more continuous nature of the commodities the consumer purchases. While the consumer may not buy three and one-half packages of some commodity, one generally thinks of the commodities purchased as being continuous. Especially when the items bought are measured by weight or volume, the discreteness imposed by fixed package sizes can safely be ignored. On the other hand, the discreteness of its available

projects is an essential part of the enterprise's capital-allocation problem. (Of course, when one considers consumer durables, the consumer's problem comes even closer to the firm's capital-investment problem as discrete alternatives become more prevalent.) The all-or-nothing basis upon which many investment projects must be accepted or rejected also means that if the number of commodities considered in the consumer problem is reasonably limited, corner solutions are to be expected much more often in the capital-budgeting case than in the constrained-consumer one.

Nevertheless, with each having an objective function to be optimized -- and it shall be seen later that these two criterion functions are more closely related to one another than one might think at this point -- and budget constraints subject to which the decisions are made, the two problems are analogous. This is especially true if one considers a consumer making plans over a period of time so that several consumption periods and hence several budgets must be considered. While interrelationships of the mutual-exclusion type might be hard to find, there may well exist commodities that make up contingent pairs for the consumer. Insofar as supplementary restrictions are concerned, one is reminded of the literature on simple and point rationing of consumer commodities.<sup>18</sup> Simple commodity rationing is also analogous to the absolute borrowing limits imposed on the enterprise. Lastly, both problems contain nonnegativity restrictions. The firm cannot undertake a negative investment and the consumer cannot purchase a negative quantity of any commodity. The basis for the analogy between the consumer's constrained-maximization problem and the firm's capital-budgeting problem rests, then, with the idea of utility-function maximization subject to budget constraints and supplementary restrictions.

The second situation to which the programming model of the capital-budgeting problem is analogous has already been mentioned, namely, the portfolio problem of individual or institutional investors. In the portfolio problem the

---

<sup>18</sup> See, for example, Samuelson (1947, pp. 163-171).

goal is the optimal allocation of a fixed amount of funds among financial assets while in the capital-investment problem the aim is to select the best set of physical assets (including such things as research and development programs) to purchase within given budget limitations. Since the common theoretical approach to the problem of portfolio selection, based on H. Markowitz's work,<sup>19</sup> will be relevant to the discussion of capital budgeting in a risk environment that appears later, it seems best to postpone an extended discussion of the individual's investment problem until that time.

Let us note, though, a few of Markowitz's general remarks that are analogous to those made earlier about the capital-budgeting problem. He writes:

- A portfolio analysis is characterized by
- (1) the information concerning securities upon which it is based;
  - (2) the criteria for better and worse portfolios which set the objectives of the analysis; and
  - (3) the computing procedures by which portfolios meeting the criteria in (2) are derived from the inputs in (1).<sup>20</sup>

The "information" in (1) is contained in the objective function and the constraint set spoken of earlier while the "criteria" referred to in (2) constitute the yet-to-be-defined objective function of the capital-budgeting problem. While no specific "computing procedures" have yet been mentioned for the firm's problem, it is clear that they will come from the realm of mathematical programming just as the critical-line method developed by Markowitz for the portfolio-selection problem belongs to that set of techniques.

Just as the discussion in this study emphasizes the need to consider groups of investment projects as opposed to examining the possibilities one by one in the capital-budgeting problem, Markowitz makes the same point with respect to securities.

<sup>19</sup>Markowitz (1959).

<sup>20</sup>Ibid., p. 205.

A good portfolio analysis is more than a long list of good stocks and bonds. It is a balanced whole, providing the investor with protections and opportunities with respect to a wide range of contingencies. The investor should build toward an integrated portfolio which best suits his needs....

A portfolio analysis starts with information concerning individual securities. It ends with conclusions concerning portfolios as a whole. The purpose of the analysis is to find portfolios which best meet the objectives of the investor.<sup>21</sup>

Later on in his book, he reiterates the same "basic principle" that "One must think of selecting a portfolio as a whole, not securities per se."<sup>22</sup> The need to consider sets of assets rather than individual assets separately is thus a salient common feature of both types of investment problem.

The final type of problem with which the programming model of capital budgeting can profitably be compared is the standard activity-analysis formulation of the production problem. Aside from the fact that discrete, indivisible alternatives appear much more frequently in the capital-budgeting case, the two problems are very closely related. In each instance one has restrictions subject to which the firm must operate. Financial constraints will appear more often in the capital-budgeting model than in the standard production one while physical-resource ceilings will be more prevalent in the production problem.<sup>23</sup> Nevertheless, in both settings there is the goal of optimizing the firm's objective function subject to the conditions imposed by the physical and financial environment in which the firm operates.

Most important in the analogy is the fact that in both cases, guided by its objective function, the firm seeks the optimal combination of activities from among the set of efficient combinations of activities. In the production problem each production activity is defined by a vector. Similarly, in the capital-budgeting problem, each project can be described by a vector showing its cash inflows and

<sup>21</sup>Ibid., p. 3. . Italics are mine.

<sup>22</sup>Ibid., p. 114.

<sup>23</sup>Recall, however, the comment made earlier about the frequency and importance of material restrictions in capital-budgeting situations.

outflows and any demands it places on other firm resources as well as any role it plays in interrelationships among the proposed investments.<sup>24</sup> The method of approaching the two problems is, then, formally identical. In each case, one seeks the combination of these vectors that optimizes the given criterion function.<sup>25</sup>

### 3.6. Risk and the Capital-Budgeting Decision

One task remains before a more detailed discussion and analysis of the capital-budgeting model presented here can take place. The enterprise's knowledge of the environment in which it exists, particularly with regard to the future inflows and outflows of the projects it considers, must be specified. Taking into account the decisionmaking unit's inability to see into the future with any real degree of clarity, the firm's capital-budgeting decision is assumed to be made in a risk environment.

The environment is intentionally described as one of risk rather than uncertainty. In distinguishing between these two types of nondeterministic setting, the original line of demarcation drawn by F. H. Knight is here somewhat modified. Knight wrote that

The practical difference between the two categories, risk and uncertainty, is that in the former the distribution of the outcome in a group of instances is known (either through calculation a priori or from statistics of past experience), while in the case of uncertainty this is not true, the reason being in general that it is impossible to form a group of instances because the situation is in a high degree unique.<sup>26</sup>

This distinction is based upon the frequency approach to probability: there either does or does not exist an objective probability distribution over the outcomes of

---

<sup>24</sup>The vector description of projects in the capital-budgeting problem will be explicitly given and discussed further in Chapter 5.

<sup>25</sup>For a discussion of this view of the production problem see, for example, Koopmans (1951b).

<sup>26</sup>Knight (1921, p. 233). See Van Moeseke (1965) for a point of view analogous to the one presented in this study and Farrar (1962) for a slightly different approach to the risk-versus-uncertainty distinction.

the stochastic variable. If such a distribution exists, then the environment is one of risk; if one does not exist, the situation is described as uncertainty.

L. J. Savage has shown, however, that if an individual, making a decision in an environment which Knight's distinction would have us call uncertainty, behaves with a consistency described by a particular set of (reasonable) axioms, then he acts as if he attached his own personal or subjective probabilities to each possibility.<sup>27</sup> Moreover, it can be shown: 1) that when objective probabilities exist with regard to the possible occurrences, these personal probabilities equal the objective ones; 2) that the personal probabilities can mix on equal footing with objective probabilities when some objective information is at hand; and 3) that subjective probabilities obey the same algebra as objective probabilities.<sup>28</sup> Adherents to the personal probability approach would argue, in fact, that probabilities are always subjective. The only meaning one can attach to the notion of objective probabilities, they would say, is that many -- if not all -- people hold the same subjective probability distribution over the possibilities in a given situation.

Whether one accepts the position that only personal probabilities exist or advocates the existence of both subjective and objective probabilities seems to be of little importance for the present discussion. It does, however, seem imperative to recognize that personal or subjective probability distributions do exist and are used by individuals (or corporations) in making decisions. For the theorist analyzing such decisions to treat such distributions differently than he treats objective probability distributions when they are used in the decisionmaking process seems unreasonable. When a probability distribution over the possible outcomes is invoked in a decisionmaking situation, it is employed in the same way whether it is

---

<sup>27</sup>Savage (1954, Chapters 3 and 4).

<sup>28</sup>For a brief indication of proofs of these results, see Markowitz (1959, pp. 267-269).



objectively or subjectively derived. Moreover, some would argue that "Even if he [the decisionmaker] adopts what he considers to be an objective probability distribution, the very act of adoption makes the distribution subjective."<sup>29</sup>

The distinction that seems more useful, then, is that between situations in which decisions are made with knowledge of the relevant probability distributions -- be they objective or subjective -- and those in which the decisionmaker does not use such probability distributions. The former I will refer to as risk, the latter as uncertainty. Observing this demarcation, the capital-budgeting problem is here considered as a decision made in a risk environment.

The capital-budgeting decisionmakers have an idea of the probabilities with which different net returns will be realized on projects they consider. Neither the objective or subjective nature of the probability distributions nor the precision of the decisionmaker's estimate of those distributions is at issue. It is also not to be inferred that the author believes such distributions are arrived at easily. To temper the skeptical reader's pessimism, however, let us note the experience of one consulting economist, D. B. Hertz. He writes

It has been our experience that for major capital proposals managements usually make a significant investment in time and funds to pinpoint information about the relevant factors. An objective analysis of the values to be assigned to each can, with little additional effort, yield a subjective probability distribution.<sup>30</sup>

Clearly, for a venture being undertaken for the first time, no objective probability distribution is possible. Ideas of success or failure or degree of success or failure for such a project can only be described in subjective terms. For other proposals, however, past experience may provide a reliable basis for objective probability distributions describing the returns and outlays on the projects.

---

<sup>29</sup>Hakansson (1966, p. 30, fn. 1).

<sup>30</sup>Hertz (1964, p. 100).

The main point is that major investment planning is properly treated as a problem in decisionmaking under risk, as the latter has been defined above. Those making the capital-budgeting decisions are assumed to make their plans with an idea of the probabilities with which different values of the stochastic variables will be realized. The concern of the present study will be with the case in which the gross returns on proposed projects are stochastic while the gross cash outlays on these projects are known with certainty. The chapters that follow will explain the reason for this asymmetric treatment of returns and outlays, and it will be shown how a period-analytic view of the firm's cash-flow operations can justify this treatment.

### 3.7. Two Disclaimers and a Final Note at the Outset

The introductory view of the capital-budgeting problem with which this study is concerned is now complete. Two disclaimers are, however, in order before we continue. First, no attempt is made in this study to consider the firm's capital-budgeting decision as part of a general-equilibrium model. There will be no attempt to integrate the portfolio selection of individual and institutional investors and the capital-project selection of corporations. This is not to imply that individuals want the firms in which they own stock to make all the transfers of their funds through time. On the contrary, the individuals and institutions will themselves alter their own time-streams of funds via changes in their portfolios. But I shall not take account of their portfolio changes in what follows. Instead of taking the general-equilibrium approach, which is undoubtedly more desirable in the long run, and treating the firm's capital-budgeting decision as a decision made in a social context affected by the simultaneous decisions of other individuals and firms, I shall treat its capital-budgeting problem as a "game against nature". The data of the problem, including future outlays and returns or their probability distributions, are taken as given and not as simultaneously determined by the decisions of other persons or other firms.

Second, it would be desirable to consider the capital-budgeting question as a problem in sequential decisionmaking. Indeed, this too was one of the extensions suggested by Weingartner at the end of his book. There he called for "an exploration of the dynamic properties of an iterative decision procedure in which models of the type formulated [in his book] are solved periodically, each time with more recent and hence better estimates of future flows resulting from current investments and more complete statement of future investment alternatives."<sup>31</sup> This extension, although highly desirable, will also remain for others to pursue at the close of the present study. While it is clear that a complete description of a decisionmaker's situation should include his expectations about his future post-horizon resources and uses for funds, this study will curtail his purview at the end of the horizon period. We are concerned with a firm's selection of a capital-investment program for a T-period horizon, including all opportunities that may arise within the T periods and none that may arise beyond the Tth period.

Having stated these disclaimers, the more detailed discussion of the capital-budgeting problem under risk can now begin. With the aid of mathematical programming it is hoped some valuable insights into the problem can be contributed. The aim is to develop a systematic approach to capital budgeting under risk and an idea of how optimal decision procedures might be framed. It seems fitting to close this chapter with some words of wisdom written by a well-known student of capital budgeting, George Terborgh.

I have no illusion that an improved MAPI formula [the formula for which he is responsible] -- or, indeed, any other -- can solve the capital-investment problem of management. Many of these problems are so complex they defy formula solutions. Nevertheless, a good formula [or formulation of the problem], intelligently applied, can be a valuable aid to judgment.<sup>32</sup>

---

<sup>31</sup>Weingartner (1963, pp. 193-194).

<sup>32</sup>Terborgh (1956, p. 309).

## CHAPTER 4

### THE CAPITAL-BUDGETING PROBLEM UNDER CERTAINTY:

#### THE GOAL OF THE INVESTMENT PROGRAM

#### AND PREVIOUS SUGGESTIONS FOR ACHIEVING IT

### 4.1. Introduction

The main concern of this study is, as its title and the previous chapters have indicated, the problem of capital budgeting under risk. Chapter 2 discussed the general problem of attitudes toward risk in multiperiod situations. In the previous chapter an introductory look was taken at the capital-budgeting problem and at the particular form in which it shall be considered in this study. A risk environment is the one in which corporations make their investment decisions. Nevertheless, a logical first step in approaching the question of capital budgeting in such surroundings is to consider the simpler matter of project selection in a world of certainty. In the next two chapters, we turn to a consideration of capital budgeting under certainty both for its own sake and for the insights it provides into the problem of capital budgeting under risk.

More specifically, the principal goal of the next two chapters is to develop a programming model of the capital-budgeting problem under certainty when the imperfect capital market described in Chapter 3 exists. The first writers to bring programming tools to bear on the capital-budgeting problem were A. Charnes, W. W. Cooper and M. H. Miller.<sup>1</sup> It was left to H. M. Weingartner<sup>2</sup> to establish firmly the power of programming methods in dealing with the question of capital budgeting. What follows in the next chapters is another contribution to the programming literature on capital budgeting that has followed the path broken by these writers.

---

<sup>1</sup>Charnes, Cooper, and Miller (1959).

<sup>2</sup>Weingartner (1963).

Consideration is first given to the question of the goal of the capital-investment program. After arguing for a particular formulation of this objective, the present chapter will consider the rules and methods discussed in the literature -- including the previous programming approaches -- in terms of their success in achieving this goal. Since the writings on these rules and methods are readily available and have been quite adequately summarized in previous literature, no full-scale review of them will be provided here.<sup>3</sup> Instead, the discussion in this chapter will seek to evaluate and to compare these prior suggestions in terms of the goal set out here. The following chapter will continue the discussion by reviewing some previous findings on the actual budgeting of capital by "real-world" firms. These findings will serve to point up the extremely normative role of this study and, in fact, of most theoretical writing on capital budgeting. The remainder of Chapter 5 presents this study's programming model of capital budgeting in an environment of certainty.

#### 4.2. The Goal of The Capital-Investment Program

##### 4.2.A. The Traditional Goal and Its Problems

The goal of the capital investment program that a firm undertakes is clearly to improve the position of some individual or group of individuals. The objective function for the capital-budgeting problem will depend on the returns from the various alternative investment proposals and upon the set of accepted investment proposals. The important questions are whose position is to be improved and how is the improvement to be assessed, and how does the criterion function to be optimized depend on the individual projects?

The overwhelming majority of the capital-budgeting literature, especially that portion written by people associated with business schools, argues that the goal of a firm's capital program is to maximize the value of the owners' equity.

<sup>3</sup> Rather than begin to enumerate at this point the places where this literature is available, the reader is referred to the citations later in this chapter and the listings in the Bibliography.

Alternatively, the capital program -- we are told -- aims at maximizing the value of the owners' income from the business.<sup>4</sup> For example, one reads

It is conceivable, of course, that financial policy might indeed be aimed at something other than maximizing the value of the stockholders' capital. It is hard, however, to see what the other something would be. Even a corporate management completely removed from ownership and concerned only with maximizing their personal net worth (e.g., by increasing salaries) would appear to be in the best position to do this if they made investment and financing decisions in a way designed to maximize the present value of future cash streams.<sup>5</sup>

As E. Solomon, editor of The Management of Corporate Capital, writes, "All of the essays in this volume assume that the goal of capital management is the maximization of long-run earnings to present stockholders. This is not to deny that companies may have other goals, e.g., size, product line, market share, production methods, or employee relations -- but each of these specific sub-goals is a means to an end rather than an end in itself."<sup>6</sup>

This goal of maximizing the long-term earnings of present stockholders is then translated into the goal of maximizing the discounted present value of the stream of future net cash flows into the firm from the investment program and from its other resources. The statement by Roberts quoted above indicates that these two views of the aim of the capital program are equivalent for some capital-budgeting theorists. As two other leading students of firm investment decisions, J. H. Lorie and L. J. Savage, have written, "Assume that the firm's objective is to maximize the value of its net worth ... as measured by the present value of its expected cash flows. This assumption ... is equivalent to asserting that the corporate management's objective is to maximize the value of the owner's equity or, alternatively, the value of the owner's income from the business."<sup>7</sup> Weingartner

---

<sup>4</sup>For example, see Durand (1952, pp. 92-93); Gordon and Shapiro (1956, p. 142); Kuh (1960, pp. 64, 66, 78); Lintner (1965, pp. 28-29); Lorie and Savage (1955, p. 57); Roberts (1957, p. 199); Solomon (1959, p. 13); and Weingartner (1963, Chapters 8,9).

<sup>5</sup>Roberts (1957, p. 199).

<sup>6</sup>Solomon (1959, p. 13).

<sup>7</sup>Lorie and Savage (1955, p. 57).

also asserts that if the firm is taken "to maximize the net value of assets, financial and physical, as of the horizon, where the former are expressed in terms of the funds available for 'lending' at that time, and the latter are represented by the discounted streams of net revenues past the horizon,"<sup>8</sup> the model describing its actions is also equivalent to one in which the present value of the firm is maximized.<sup>9</sup>

In short, the overwhelming majority of writers in the field of capital budgeting have taken as the goal of the firm's investment program

$$(4.1) \quad \text{Maximize} \quad \sum_{t=1}^T D_t V_t$$

where  $V_t$  is the net cash inflow into the firm in period  $t$ . The symbol  $D_t$  denotes the factor applied to the cash flows in period  $t$  in order to discount them to their present value. It is the discount factor by which flows in the tth period must be multiplied if they are to be put on an equal footing with flows in the present period. In particular, then,  $D_1$  equals unity.

The important question concerning the objective given in (4.1) is "What should the discount factor  $D_t$  be?" Clearly, it is to measure the opportunity cost of postponing receipt of the marginal dollar in the tth period to that period rather than getting it now. The ratio  $\frac{D_t}{D_{t-1}}$  should equal the marginal rate of substitution of dollars in period  $t$  for dollars in period  $t-1$ . Alternatively, the ratio  $\frac{D_{t-1}}{D_t}$  should equal the opportunity cost of postponing receipt of the marginal dollar from period  $t-1$  to period  $t$ .

The literature, by and large, tells us to discount using the company's "cost of capital" as the discount rate. That is, set  $D_t$  equal to the t-1st power of the reciprocal of one plus the company's "cost of capital." If a perfect capital market exists, the cost of capital is well-defined. When unlimited amounts of money can be borrowed or loaned at constant and equal rates of interest with no

---

<sup>8</sup>Weingartner (1963, p. 141).

<sup>9</sup>Ibid., p. 142. While his conclusion is correct, Weingartner's alleged proof does not seem acceptable. An alternative proof is given in footnote 84.

transactions costs, the cost of capital is the interest rate, call it  $r$ ,<sup>10</sup> that prevails in the market. The firm borrowing \$1 in period  $t-1$ , must pay back  $\$(1+r)$  in period  $t$ . Similarly, if the firm invests \$1 it has in period  $t-1$  in a particular physical project it foregoes the net return of  $\$r$  in period  $t$  that it could have obtained by lending in the market. The opportunity cost of postponing receipt of one dollar for one period, that is,  $\frac{D_{t-1}}{D_t}$ , is  $\$(1+r)$ . In short, \$1 in period  $t-1$  is equivalent to  $\$(1+r)$  in period  $t$ . Hence, to reduce  $V_t$  to period  $t-1$  dollars one would divide  $V_t$  by  $1+r$ . Continuing this recursion back to the present, one finds  $D_t$  in (4.1) equals  $(\frac{1}{1+r})^{t-1}$ .<sup>11</sup>

Once one leaves the world of perfect capital markets, though, the "cost of capital" loses much of its clarity as an economic concept. Some leading capital-budgeting theorists simply circumvent the issue. Lorie and Savage, for example, write, "The question of determining the cost of capital is difficult, and we, happily, shall not discuss it. Although there may be disagreement about methods of calculating a firm's cost of capital, there is substantial agreement that the cost of capital is the rate at which a firm should discount future cash flows in order to determine their present value."<sup>12</sup> Weingartner simply states that he agrees with Lorie and Savage that future cash flows should be discounted at the company's cost of capital. He goes on to say "how this concept is to be measured is not discussed by them [Lorie and Savage], and will not be discussed by us."<sup>13</sup> More recently, in a paper that emphasizes the presence of risk and project inter-relationships in capital budgeting, F. S. Hillier writes of his discount rate -- the cost of capital -- "It is assumed throughout this paper that ... [it] is known.

<sup>10</sup> Recall that it is assumed that there are intertemporal differences neither in the borrowing rate nor in the lending rate.

<sup>11</sup> If the rate of interest could vary from one period to the next, denoting the perfect capital market's interest rate in the  $i$ th period by  $r_i$ , one finds

$$D_t = \prod_{i=1}^{t-1} \left( \frac{1}{1+r_i} \right) .$$

<sup>12</sup> Lorie and Savage (1955, p. 57).

<sup>13</sup> Weingartner (1963, p. 8).



However, the precise determination of its proper value is a deep theoretical question which continues to receive serious study elsewhere."<sup>14</sup>

The exact determination of a company's cost of capital has, indeed, continued to be a subject of active discussion in the capital-budgeting literature. Recently, for example, M. H. Miller and F. Modigliani have presented new methods for inferring a company's cost of capital from data on the market value of its securities. In particular, they present empirical estimates of this elusive figure for a sample of large firms in the electric utility industry.<sup>15</sup> Nor has this question of determining a firm's cost of capital been the concern of capital-budgeting theorists alone. For example, a member of the management of a large oil company described the situation of investment planners in his organization as follows.

There is one major theoretical and practical problem in using the discounted-cash-flow procedure [internal-rate-of-return method] for which we have not yet found a fully satisfactory solution. This problem is that of developing a return-on-investment figure for whole departments or groups of departments which may be computed year by year and compared with the returns calculated under the discounted-cash-flow procedures at the time individual projects were undertaken.<sup>16</sup>

The cut-off rate for which McLean and his associates were groping is, in fact, the company's cost of capital.

Unfortunately, despite the serious study the question has received, there remains much disagreement about exactly how the cost-of-capital figure should be determined. But there does remain substantial agreement that it is the figure to be used in discounting future cash flows to the present. A British writer has described the situation surrounding the cost of capital as the appropriate discount rate all too well.

---

<sup>14</sup>Hillier (1964, p. 7).

<sup>15</sup>Miller and Modigliani (1966). A recent theoretical paper by Baumol and Malkiel also sheds some new light on theoretical aspects of the measurement problem. Baumol and Malkiel (1967).

<sup>16</sup>McLean (1958, p. 69).

Most economists who have considered this question [of the appropriate discount rate] have stated that the appropriate rate to use is that which the firm has to pay for capital, i.e., the 'cost of capital.' The present author, however, regards the 'cost of capital' as one of those quasi-mythological creatures, like the Loch Ness Monster -- everybody has heard of it; a few people claim to have seen it, but nobody has ever run it to earth.<sup>17</sup>

Even in a model where one is not confronted with the great variety of securities that exists in reality, discounting at the company's cost of capital is not a straightforward procedure. For example, what is the appropriate discount rate when the firm can neither borrow nor lend funds? This question of investment decisionmaking in an environment of pure capital rationing has attracted considerable interest in the theoretical literature.<sup>18</sup> Alternatively, consider an imperfect capital market of the type described in the previous chapter. Such a model has generated much controversy in the literature.<sup>19</sup> Should future cash flows be discounted at the borrowing rate or the lending rate when the two diverge? In the presence of an imperfect capital market, the determination of each  $D_t$  in equation (4.1) and the entire discounting procedure is not so clear-cut as it is when a perfect capital market exists.

The problem is that for a firm having no recourse to a perfect capital market, the relevant discount rate is not an independent entity. Rather it is itself a product of the analysis that determines the firm's optimal investment program. The discount rate cannot be arbitrarily assigned the value of the borrowing rate or the lending rate in the case where the two diverge. In fact, as Hirshleifer has shown so neatly, the appropriate marginal discount factor  $\frac{D_t}{D_{t-1}}$  may be a function of neither the borrowing rate nor the lending rate. This would be the case in which the firm neither borrows nor lends in its optimal

---

<sup>17</sup>Adelson (1965, p. 38).

<sup>18</sup>Hirshleifer (1958); Baumol and Quandt (1965); Manne (1966).

<sup>19</sup>Dean (1951); Lutz and Lutz (1951); Renshaw (1957); Roberts (1957).

position in the  $t$ th period. In this instance the appropriate discount rate is derived from the productive investment opportunities within the firm.

Consider a two-period example following Hirshleifer's type of analysis.<sup>20</sup> On the horizontal axis in Figure 4.1 we represent a decisionmaker's actual or potential income in period  $t-1$ ,  $W_{t-1}$ , while on the vertical axis actual or potential income in period  $t$ ,  $W_t$ , is shown. There is an indifference map showing preferences among combinations of income in period  $t-1$  and period  $t$ . The curves labeled  $U_1$  and  $U_2$  in Figure 4.1 are two of the decisionmaker's indifference curves.

The investment opportunities available are of two types: production opportunities that involve real productive transfers of income through time (physical investment) and market opportunities that involve transfers through borrowing and lending. In Figure 4.1, one of the individual's or firm's production-opportunities locus is shown as curve  $PP'$ . A market-opportunities locus for borrowing is  $BB'$  while that for lending is  $LL'$ . It should be noted that the absolute slope of  $BB'$  is greater than that of  $LL'$ . This is so because the absolute slope of  $BB'$  is unity plus the borrowing rate ( $1+r_B$ ) and the absolute slope of  $LL'$  is unity plus the lending rate ( $1+r_L$ ). In the divergent-rates case it is assumed that  $r_B$  is greater than  $r_L$  for otherwise if  $r_L > r_B$  it would be possible to earn an infinite amount of money by borrowing and simultaneously lending these borrowed funds.

The objective is to move to the highest possible indifference curve given the opportunities available -- both productive and market opportunities. In Figure 4.1, the highest indifference curve attainable is  $U_1$  and the investor's equilibrium is at  $E$ . But here he is neither borrowing nor lending! The opportunity cost of capital is neither the borrowing rate nor the lending rate. It is, instead, some rate in between  $r_L$  and  $r_B$  and is equal to the absolute slope minus unity of  $U_1$  and  $PP'$  at the point of tangency  $E$ .

---

<sup>20</sup>Hirshleifer (1958, pp. 207-212) is the model for this example.

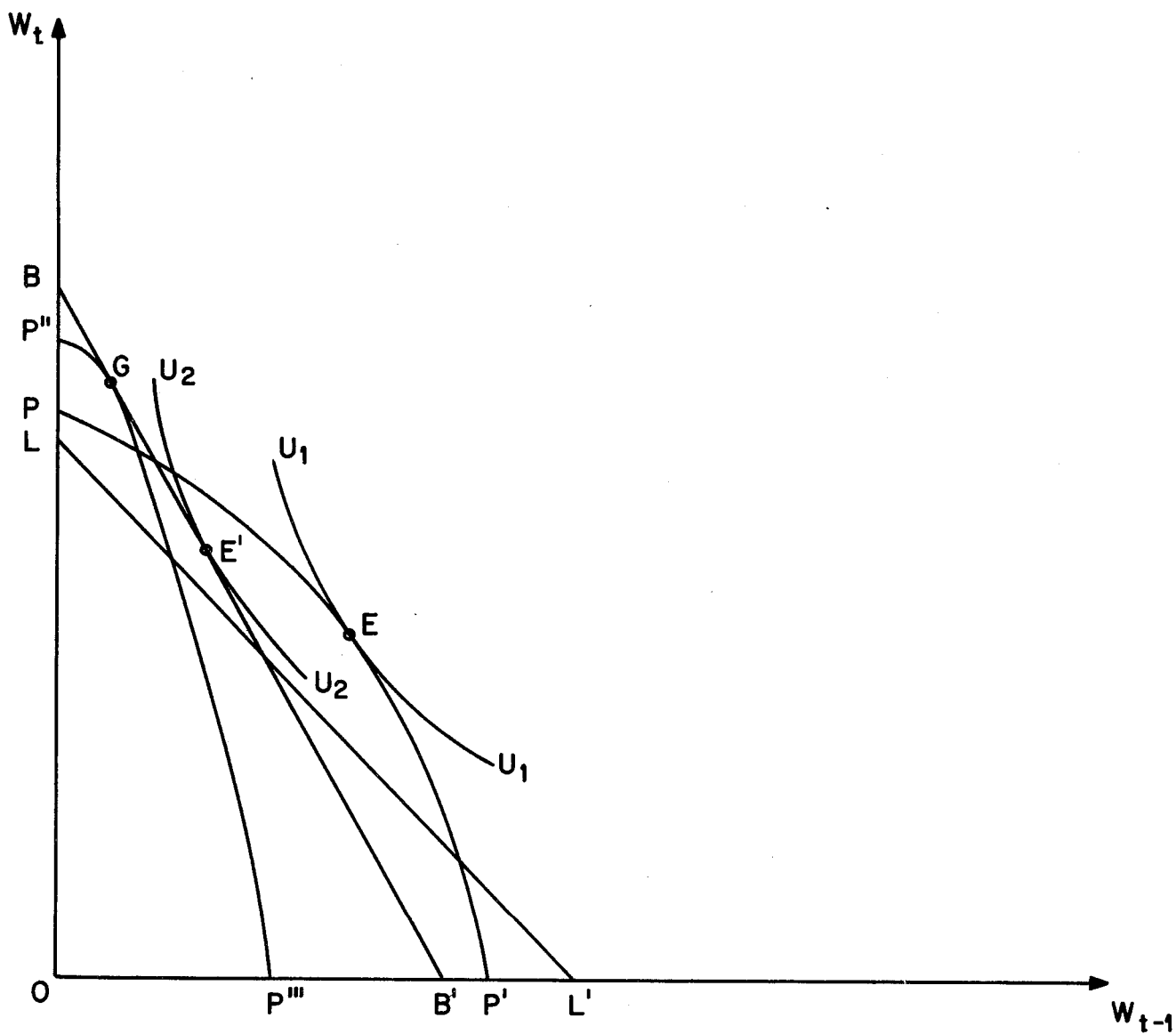


FIGURE 4.1

The Case of Divergent Borrowing and Lending Rates  
with  
(1) Neither Rate the Correct Discount Rate  
(2) The Borrowing Rate the Appropriate Discount Rate

Alternatively, suppose the production-opportunity locus were  $P''P'''$ . Then the highest indifference curve attainable would be  $U_2$  with equilibrium occurring at  $E'$ . Note that  $E'$  is attained in two steps. First, the decision-maker pursues the productive opportunities to the point of tangency,  $G$ , between the production-opportunities locus and the borrowing-opportunities curve. Then he borrows backward along  $BB'$  until he reaches equilibrium at  $E'$ . Again the appropriate discount rate is derived from the slopes of the opportunity locus and the indifference curve at equilibrium. But here this slope is  $-(1+r_B)$  and hence the appropriate cost of capital is the borrowing rate. A similar case could be given where  $r_L$  is the correct discount rate.

The important point the analysis illustrated in Figure 4.1 helps bring to the fore is the interdependence involved in determining the optimal investment program and the correct discount rate. If the optimum entails borrowing in period  $t-1$ , then  $\frac{D_t}{D_{t-1}} = \frac{1}{1+r_B}$ , while if it calls for lending in that period  $\frac{D_t}{D_{t-1}} = \frac{1}{1+r_L}$ . If, in contrast, the decisionmaking unit neither borrows nor lends in its optimal position for period  $t-1$ , then the appropriate marginal discount factor is the reciprocal of the absolute slope of the indifference curve and production-opportunity locus at their point of tangency. The ratio  $\frac{D_t}{D_{t-1}}$  would then equal the reciprocal of unity plus the rate of return on the best alternative project available at the equilibrium point. In any case, the correct discount factor is only known ex post, not ex ante, and hence it cannot be used in helping to determine which projects to accept and how much to invest.

#### 4.2.B. A Utility-Maximizing Approach

A discounted-present-value objective function is thus an inappropriate one for use in choosing the optimal investment program of a firm operating in an imperfect capital market. Discounting in such an environment is an artificial procedure since the appropriate discount rate is only known after the optimum

is found. Moreover, discounting at arbitrarily chosen market rates of interest can lead to results contrary to the best interests of the firm. The only correct approach to the problem of budgeting capital when the capital market is imperfect is to treat the problem as part of the general theory of choice. The goal is to maximize utility subject to the production and capital-market opportunities available and the constraints facing the firm.

This is the path followed by Hirshleifer who attempted somewhat unsuccessfully to bring the literature back to the Fisherian tradition. In his important 1958 article he wrote, "More recent works on investment decisions ... suffer from the neglect of Fisher's great contributions -- the attainment of an optimum through balancing consumption alternatives over time and the clear distinction between production opportunities and exchange opportunities."<sup>21</sup> More recently, constrained utility maximization was also the formulation at which Baumol and Quandt arrived after demonstrating the futility of attempting a present-value formulation for the case of pure capital rationing.<sup>22</sup>

Having set a utility function at the center of the correct approach to the capital-budgeting problem as it is considered here, two important questions naturally arise. First, whose utility function is being maximized, and second, what are the arguments of the utility function? The answers to these two questions are clearly interrelated. Hence, though the discussion begins by responding to the first of them, answering it completely will also entail answering the second question.

#### 4.2.C. Whose Utility Is Being Maximized?

The process of investment is not an end in itself. A rational investor does not pursue investment projects for the sheer pleasure of undertaking projects.

---

<sup>21</sup>Ibid., p. 205.

<sup>22</sup>Baumol and Quandt (1965).

Instead, the capital program must yield some "gain" to the investor, specifically it must increase his ability to consume at some future time. Together with Baumol and Quandt, then, the present work takes as the investor's goal the desire "to maximize the stream of purchasing power provided by his bundle of investments."<sup>23</sup> The utility function to be maximized is a function of the period-by-period consumption incomes derived from the investment program. And these arguments of the function are the consumption incomes of the owners of the firm -- the equity stockholders.

But how is this expression of owners' time preferences, which constitutes the objective function for the capital-budgeting problem, to be obtained? Aggregation of the utility functions of individual shareholders introduces the entire set of aggregation problems. It also brings with it the difficulties that trouble the use of social welfare functions and social indifference curves. In fact, the management making the capital-budgeting decision is not going to attempt to seek out the utility function of each individual shareowner and then suitably aggregate them. To pursue this approach would, for one thing, entail altering the maximand each time ownership of the corporation changed, even down to the level of a purchase of a single share by a new stockholder or by a present shareholder from another present owner!

Instead, the utility function to be maximized is taken to be management's perception of the owners' utility. This does not really remove the aggregation problem. The ownership of the firm may change from one period to another via share transfers and management will want to take account of these changes. There remains, then, the problem posed by Modigliani and Miller, "How can the economist build a meaningful investment function in the face of the fact that any given investment opportunity might or might not be worth exploiting depending on precisely who happen to be the owners of the firm at the moment?"<sup>24</sup>

---

<sup>23</sup>Ibid., p. 326.

<sup>24</sup>Modigliani and Miller (1958; p. 152).

The position taken here is that management attempts to predict any important changes it foresees in stockholder composition. These predictions are then embodied in the specification of the utility function, in the parameters associated with different periods' consumption alternatives. For the management of large well-established corporations, the task of prediction may be made somewhat easier by the fact that such corporations tend to attract a particular clientele and to hold it for considerable periods of time. To quote H. Bierman, Jr. and S. Smidt, two leading students of capital budgeting, for support on this point, one reads in their work:

If the stock of corporations were distributed between investors in some random manner, the conflicts of interest between investors would be of great practical importance. Actually, however, most large corporations follow reasonably consistent financial and investment policies. The securities of most listed corporations undoubtedly tend to flow into the portfolios of investors whose personal investment goals are consistent with the known policies of the companies whose stock they hold.<sup>25</sup>

It is important to emphasize at this time that it is the imperfect nature of the capital market that forces the firm to take explicit account of the ownership's time preferences. As Hirshleifer has shown,<sup>26</sup> in a world of perfect capital markets where both individuals and corporations can borrow and lend at fixed and equal rates of interest, individual utility maximization can be viewed as a two-stage process. In the first step, the well-defined discounted present value of the individual's time stream of wealth is maximized by making productive investments. Then, by borrowing or lending at the single market rate of interest, the individual can select from among the time streams of maximal discounted present value that particular combination of present and future incomes that best satisfies his time preferences. In the case where the corporation and its shareholders face a perfect capital market, then, the firm can confine its attention to maximizing

---

<sup>25</sup>Bierman and Smidt (1966, p. 155).

<sup>26</sup>Hirshleifer (1958, pp. 206-209).



the discounted present value -- calculated at the unique market rate of interest -- of its cash flows. It can leave to each individual investor the job of moving his own funds through time to obtain from this maximum wealth the bundle of present and future incomes he prefers.

This is not, however, the world of the present model. In the model being developed in this study, the imperfect capital market that exists prevents the firm from taking such a restricted view of its problem. In the presence of market imperfections, as transactions costs, that drive a wedge between the borrowing and lending rates the corporation and its shareholders face, the firm must take explicit account of its owners' time preferences. It must take upon itself more directly the task of helping its owners achieve their optimal time streams.

To dispel the cloud of naïveté which may seem to surround the previous statements, it should be made clear that there is no implication that management's task -- as just stated -- is an easy one. The job of ascertaining the composition of stockholders with regard to time preference and balancing these preferences in the form of a multiperiod utility function is no mean task. Nevertheless, it seems -- at least in a normative sense -- a duty management must perform in the imperfect capital-market setting of the present model.

This brings us to a most important point. Does management really act in the best interests of the owners? Do businessmen act in accord with the basic assumption made here, namely, that they seek to maximize some function of the owners' well-being? E. Kuh writes the following in answering a similar question.

Indeed, instances can be cited where, quite to the contrary, stockholders have been milked dry. Less flagrant situations exist, where it is hard to believe that different financing arrangements would not have been more beneficial to existing stockholders. Part of such observed disparities are due to faulty estimation of market preferences. Another part to conflicts in goals between managers and stockholders, and a third part to inadequate knowledge of available opportunities. Nevertheless the often repeated and apparently real concern voiced by business executives about

common stock values lends minimum support to the assumption as a plausible behavioral hypothesis. As a normative proposition, I believe, no serious dispute seems warranted.<sup>27</sup>

The stand taken here -- that management's perception of the owners' utility function is the maximand -- is not simply a compromise. Management does, in fact, make the decision. But its decision aims to optimize the owners' position. If nothing more motivates the managers to do so, there is at least the fact that although the owners do not continually review the actions taken by management, they do at definite points in time scrutinize these decisions.

Biases or imperfections in management's vision are not excluded. In attempting to ascertain the owners' time preferences management may err, consciously or unconsciously. The latter is simply an "error of measurement," if you will. The former, however, reflects differences in the ownership's and the management's desires and objectives. Hence, management's own ambitions may influence its perception of the corporate ownership's utility and these biases are incorporated in the utility function that emerges for maximization.

If there is to be substance to this position, the question of how different the goals of owners and the aims of managers will be must be considered. To say the objective function is an ownership utility function and then to allow for biases so different and so strong that the original maximand becomes distorted into a management utility function is not to make a very strong case. My stand is that the underlying objects of utility are the period-by-period consumption alternatives available to the owners but that when management attempts to determine the utility function some biases are introduced. The function remains, however, a repository of ownership values not of management values.

If management's aims were too drastically different from the owners' this position could not be considered a tenable one. The set of views with which

---

<sup>27</sup>Kuh (1960, p. 78).

the position taken here must come to grips is that of the behavioral theory of the firm.<sup>28</sup> In particular, it is this theory's emphasis on management's discretionary power that must be taken into consideration. Beginning with the findings of leading students<sup>29</sup> of the modern American corporation, the behavioral theory of the firm has as one point of emphasis the division between ownership and management in today's firm. This divorce has, they argue, following the lines drawn by Berle and Means, left management with much greater discretionary power in its control of the corporation. Hence, the behavioral theory continues, when discussing firm decisionmaking one must consider management's motivations and not ownership's motivations.

One proponent of this approach -- he calls them "managerial discretion models" -- suggests that the criterion function of the firm is management's utility function.<sup>30</sup> It is to be maximized subject to the constraint that reported after-tax profits exceed a certain level demanded by the owners. The arguments of the utility function to be optimized include: (1) staff size (in dollar terms), or (approximately) general administrative and selling expenses; (2) managerial emoluments, referring to that part of management salaries and perquisites that is discretionary, and including expense accounts, executive suites and so on, all essentially economic rents with zero productivities; and (3) discretionary profit, being the difference between earnings and the amount set as a minimum by the profit constraint. These pecuniary arguments are the manifestations of such nonpecuniary management motives as security, dominance (including power, status, prestige), and professional excellence, as well as manifestations of the managers' desire for salary.

Management's "expense preference" for the pecuniary items is the behavior to which these other more fundamental motives impel it. For example, staff

---

<sup>28</sup>See, for example, Cyert and March (1963) and Williamson (1963b, 1964).

<sup>29</sup>Berle and Means (1932); R. A. Gordon (1961, especially pp. xii, 305-316); Kaysen (1959, especially pp. 89-91).

<sup>30</sup>See Williamson (1963b, pp. 1032-1040; 1964, Chapters 3 and 4). What follows summarizes part of Williamson's case.

appears as an argument in the maximand as an expression of the salary motive as well as of almost all the overlapping nonpecuniary motives listed above. As Williamson tells us, "since promotional opportunities within a fixed-size firm are limited, while increased jurisdiction has the same general effect as promotion but simultaneously produces the opportunity for advance to all, the incentive to expand staff may be difficult to resist. Being a means to promotion, expansion of staff serves to advance both salary and dominance objectives simultaneously. In addition, staff can contribute to the satisfaction of security and professional achievement objectives as well."<sup>31</sup> As concerns security, "If the surest guarantee of the survival of the individual parts [of the firm] appears to be size, efforts to expand the separate staff functions can be predicted" while "The 'professional' inducement to expand staff arises from the typical view that a progressive staff is one that is continuously providing more and better services. An aggressive staff will therefore be looking for ways to expand."<sup>32</sup>

A complete reconciliation between the stand taken in this work with regard to the capital-budgeting maximand and the general position of Williamson and his colleagues is clearly not possible. But enough of a rapprochement is possible so that our position remains a reasonable one to which to hold. First, it ought to be noted that most of the behavioral-theory literature is concerned with the output decisions, the market-strategy decisions, and the internal-allocation decisions of firms. These works have not been concerned with investment decisions. There is reason to believe that in the realm of investment decisions, management will pay greater heed to the interest of the shareholders. The owners' utility functions, not the managers', will be paramount. The reason lies, in fact, within the scope of the logic of managerial discretion models.

---

<sup>31</sup>Williamson (1964, p. 34).

<sup>32</sup>Ibid.

Uppermost among management's objectives is the desire to maintain and increase the discretionary power it has accumulated. This is true for all the reasons Williamson has given. But what better way is there for management to secure its discretionary position than to pursue a continuously successful -- in the eyes of the owners -- investment program? Few items get a more impressive display in the couriers of financial news to the owners, as The Wall Street Journal or The New York Times Financial Section, than a successful capital-investment program. Meeting a minimum-profits constraint may "satisfice" the owners but carrying on an aggressive, successful capital program will encourage them to transfer even greater discretionary power to the managers.

A second important point in the relationship between the behavioral theory's approach and the position put forward earlier in this chapter works through Williamson's "discretionary profits" argument. These profits, equal to total (after-tax) reported profits minus the minimum-profit figure, are supposed to be used for discretionary investments by the managers. Investments that are necessary because of economic considerations are included in the minimum-profit constraint. Concerning the role of discretionary profits Williamson writes the following.

That the managers should desire to earn profits that exceed the acceptable level derives from the relationship that profit bears to discretion, self-fulfillment, and organizational achievement. Since expansion of staff and emoluments can scarcely proceed independently of the expansion of physical facilities and since financing of this expansion (whether from internal or external sources) will be tied to profitability of the firm, profits in excess of the minimum acceptable level may well be desired by the management. Moreover, managers derive satisfaction from self-fulfillment and organizational achievement and profit is one measure of this success.<sup>33</sup>

Insofar as the indirect effects of discretionary profits on allowing expansion of staff and emoluments are concerned, physical expansion is thus a means for

<sup>33</sup>Ibid., p. 36.

increasing discretionary power. Yet the investments undertaken must satisfy the owners in order for them to be willing to transfer even more stewardship of the firm to the managers. Hence, in exercising this discretionary power, it is in the managers' best interests to be duly attentive to the desires, in particular, the time preferences of the owners.

The owners' utility function is thus the one to be maximized in the case of the capital-budgeting decision, for it is by optimizing it that management can come closest to achieving its own goals. Of course, misperceptions may enter as management tries to specify the owners' utility function. If, for example, management is quite intent upon a rapid expansion of staff, the resulting "perceived" utility function may show greater impatience than the owners' would. Williamson's utility-function arguments not only seem to point the way to the ends management is actually trying to serve but also to sources of bias in the process of specifying the capital-budgeting objective function.

As a final point in discussing the behavioral theory of the firm and its relationship to the objective function argued for here, note that it is top management that makes the final capital-budgeting decisions. If top management does not speak the final word for individual projects, it at least makes the decision with regard to total amounts allocated to various departments and so on. There is, of course, a need to establish a certain consensus among these top executives. This is achieved largely through a screening-selection process for those who do make it to the top and through a socialization or "inbreeding" process. "The former tends to provide members to the firm with compatible characteristics, the latter is a process by which the major values of the groups are internalized by its individual members."<sup>34</sup> In any event, as R. A. Gordon has written, "the ... rule is that corporate executives as a group determine the volume and direction of investment in their firms ..."<sup>35</sup> and the motivations of these corporate executives are likely

---

<sup>34</sup>Ibid., p. 154. For a good discussion of the question of social choice and group decisionmaking within the firm, see Williamson (1964, Chapter 8).

<sup>35</sup>R. A. Gordon (1961, p. 114).

to be quite close to those of the company owners. This will be especially true in firms where management, and especially top management, has significant stock options available to it. Williamson himself in his contribution to the Cyert and March volume notes that his "argument runs in terms of 'managers' rather than 'top management'," particularly to provide for those cases "where the goals of top management might be approximately those of the stockholders."<sup>36</sup> With "top management" playing such an important role in the capital-budgeting decision one might thus expect less misperception in seeking the owners' utility function than if the verdict were to be totally in the hands of lower-level managers.

#### 4.2.D. The Arguments of the Utility Function

This concludes the case for the position that the appropriate objective function for capital budgeting is the utility function of the owners' consumption alternatives in the T periods, as that function is seen by management. It remains to answer more precisely the second question posed earlier concerning the criterion function. Specifically, the exact nature and measure of these consumption alternatives must be indicated. The value to the owners of increments to their consumption opportunities is properly measured only by considering the size of their entire store of assets and not the size of these increments alone.<sup>37</sup> Keeping track of each individual shareholder's general asset position and incorporating changes in them (by prediction or otherwise) into the utility function serving as the maximand is, however, completely unrealistic. It certainly goes even a good bit further in the direction of gross impossibility than the other assumptions made earlier about management's predictive power.

Instead, the arguments in the utility function will be taken to indicate the changes in (that is, the increments to) stockholders' consumption possibilities

<sup>36</sup>Williamson (1963a, p. 241, footnote 16).

<sup>37</sup>Borch (1965, Chapter IV, p. 32).

generated by the firm's operations, including the investment program, in the several periods of the planning horizon. The reason for using increments generated by all of the firm's operations and not by the investment program alone is to minimize the error involved in considering only increments to consumption alternatives rather than entire consumption alternatives. That is, optimally one would want to consider the size of an owner's entire consumption possibility in a period. But the firm cannot realistically do this. Hence, attention is restricted to the increments to this consumption alternative that the particular firm under consideration provides. This is at least somewhat more realistic than trying to consider the owners' entire consumption alternatives in each period. On the other hand, if attention were restricted to considering only the increment that the investment program itself provides, the true value of the program would be misrepresented even more. The procedure here at least allows for the fact that the owners may have some basic consumption possibilities stemming from the operations of this firm that exist apart from the investment program. It at least takes cognizance of the consumption alternatives resulting from the use of resources this firm actually controls prior to the start of the investment program, that is, at  $t = 0$ .

Some defense for the use of such a "rule of the perfect miser" in the present case comes from Borch who so adequately chastises people who suggest the general use of such a rule. He writes:

It has been suggested that a large corporation may need a decision rule of this kind. In such corporations decision making must necessarily be delegated to a great number of executives who cannot always be fully informed about the situation of the corporation. They will then need to have the basic policy of the corporation spelt out in decision rules which are 'fool-proof' in the sense that they cannot lead an executive astray, even if he should not be quite up-to-date on the actual situation of his corporation.<sup>38</sup>

---

<sup>38</sup> Ibid., Chapter IV, p. 33.



If, in this quotation, the phrase "situation of the corporation" is replaced by "situations of the individual shareholders," a good justification -- or should I say rationalization -- of the position presented here is provided.

The increment to a shareholder's consumption possibilities arising from the operations of the firm is some function of the dividends per share and the capital gains per share attributable to those operations. It should be clear that in a finite-horizon model it does not suffice to measure this contribution by only the dividend stream the firm provides. If the firm's performance causes an increase in the market value of the stock, the shareholder has the option of selling shares and receiving the capital gains. Whether a particular stockholder would prefer to receive earnings generated by the firm as dividends or as capital gains depends, for one thing, on the income tax rates in his tax bracket. It also depends on any subjective attitudes he might have toward selling the stock, toward the timing of the receipt of income, his other investments, and so on. Nevertheless, the increase in the owner's consumption possibility generated by the firm in question consists of dividends and potential capital gains, suitably adjusted for differing tax rates. The increase in his utility caused by this consumption-alternative increment depends on an explicit statement of his utility function.

This leaves us with two major problems. First, how can dividends and increases in the market value of the stock be analyzed into the part due to the actual performance of the firm and the part due to other causes? Numbered among these other causes are things like simply psychological or emotional reactions of the market, poor judgment of the firm's earning potential on the part of the market, the market's anticipation of earnings of investments not yet undertaken, and management's concern for its image. Second, since individuals have different preferences regarding the timing of returns and the distribution of earnings between retentions and dividends are we not forced back to having to aggregate individual owners' utility functions in order to arrive at the capital-budgeting problem's objective function?

The answer to the first question is that such an analysis is not really possible. Rather than face up to this truly unsolvable problem of resolving dividend and capital-gain streams into the portion attributable to the actual operations and the part due to other sources, I will use a proxy for the increment in the owners' potential consumption alternatives.

Denoting by  $W_t$  the true, but unknown, increment provided for the owners by the firm in the  $t$ th period, the utility function to be maximized in the capital-budgeting problem is

$$(4.2) \quad U = U(W_1, W_2, \dots, W_T) .$$

As a proxy for the true  $t$ th-period increment ( $W_t$ ) deriving from the firm's resources -- both presently owned resources and those to be acquired in the investment program -- I shall use the total amount available for withdrawal from the firm in that period. That is, the  $t$ th-period net earnings of the investment program and the  $t$ th-period net earnings of the initial firm resources will be taken as the measure of the increase in the owners'  $t$ th-period consumption alternatives.

The second question asked above can only be answered by making explicit the stand taken here on the issue of differences in individual owners' preferences regarding retained earnings versus dividends and the timing of returns. The optimal approach to the capital-budgeting decisions of firms is to treat them in a general-equilibrium setting.<sup>39</sup> In such a model, the stockholders' portfolio choices among securities of various firms and between such risky assets and cash would be determined simultaneously with the optimal capital-investment programs of the firms in question. Differences among stockholders with regard to time preferences and preferences between dividends and capital gains could be incorporated in their individual utility functions. The investments and the dividend pay-out policy

---

<sup>39</sup>Lintner (1965) goes part of the way toward such a model.

of each firm would be determined at the same time as its ownership was being determined. Clearly, from what has been said before, the objective functions of the firms would thus depend on the decisions simultaneously being taken elsewhere in the model. The utility functions of those deciding to buy stock in a firm would be suitably aggregated to arrive at the objective function for the corporate decision.

As the disclaimer in Chapter 3 stated, this general-equilibrium model is not the approach taken here. With so much still to be done in the field of capital budgeting, it seems reasonable to be content with taking but a few steps in the direction of more complex models. The aggregation problem will be dismissed by relying on the stability of the shareholding clienteles of the various corporations that was referred to earlier when discussing management's prediction problems in perceiving the owners' utility function. It is supposed, then, that previous trading in the shares of companies has caused the timing of dividends and the dividend-to-retentions ratios to be optimal for both firms and shareholders. The present model is one of partial equilibrium in which this optimality of the timing of dividends and the extent of retentions is given as having already been achieved.<sup>40</sup>

Hence, when management attempts to perceive the owners' function to maximize, it concerns itself with a function of the amounts available for withdrawal in each of the  $T$  periods. The question of the optimal timing of withdrawals is answered by the form of the utility function. The problem of differences in the desired pay-out ratio is assumed to be nonexistent as the owners of the firm are taken to be aware of and content with the firm's consistent financial policy. In short, an equilibrating process in the stock market -- and exogenous to the model being developed here -- has ensured that present and future owners will be homogeneous with respect to the desired timing of withdrawals and content with management's dividend-versus-retention policy.

---

<sup>40</sup>The extended quotation from Bierman and Smidt (1966) presented in the discussion of management's prediction problems (see footnote 25) applies once again.

#### 4.2.E. The Objective Function of the Capital-Budgeting Model

The objective function of the capital-budgeting problem is, then, the utility function shown in (4.2), where each  $W_t$  is the amount available for withdrawal from the firm in period  $t$ . This utility function can be reduced to more basic elements of the investment decision, namely, the individual projects. Unfortunately, the literature has for the most part -- the Baumol and Quandt article and Manne's paper being exceptions<sup>41</sup> -- not drawn together the consumption-alternative approach of neoclassical capital theory and the project approach of more recent capital-budgeting theory. Hirshleifer, on the one hand, never leaves his world of indifference curves and production-possibility loci among consumption alternatives. The model-builders in capital budgeting, for their part, rarely forsake their emphasis on individual projects.

These two approaches are brought together in the present model through the  $W_t$ -variables, the increments to ownership-consumption alternatives provided by the chosen set of investments and by the originally owned resources. The gross earnings of the investment program in period  $t$  plus what remains of the net earnings of the initial firm assets after the gross investment expenditures in period  $t$  have been made constitute the  $W_t$ -variable. But the returns of the investment program as a whole in the  $t$ th period are the sum of the returns in that period from the individual proposals. It must be understood in this connection that proposals will be defined in such a way as to encompass any return and outlay interactions among physical projects. Since this is a very important point, to help in effecting its understanding let us digress briefly.

Suppose that the total cash outlay required to undertake two physically distinct projects simultaneously differs from the outlay that must be made if the two projects are considered separately. For example, both projects make use of the same machine and the machine would therefore have to be bought so long as

<sup>41</sup>Baumol and Quandt (1965); Manne (1966).

either one of the two investments were made. But if both physical projects were undertaken, only one machine and not two would have to be bought. On the other hand, one might have a case in which the returns derived from undertaking two physical projects simultaneously differed from the sum of the returns if the two were to be undertaken individually.<sup>42</sup> As an example, consider the possible location(s) of a new store by a chain-store company. Building a store at location A may bring one set of revenues, building one at B might bring another set. But simultaneous construction of the two stores may lead to returns less than the sum of those from A and B separately if, for example, the two locations are sufficiently close to one another as to be in competition.

Let  $a_{ti}$  denote the gross cash inflow from project  $i$  in the  $t$ th period and  $c_{ti}$  the gross cash outlay on the project in that period. The problem of cash-outlay and return interactions may then be succinctly stated as follows. While accepting project I may require a stream of outlays  $c_{tI}$  and give rise to a stream of returns  $a_{tI}$ , and accepting project II may lead to the corresponding streams  $c_{tII}$  and  $a_{tII}$ , joint undertaking of the two may lead to the cash flows  $c_{t,I+II} \neq c_{tI} + c_{tII}$  and  $a_{t,I+II} \neq a_{tI} + a_{tII}$  for gross outlays and gross revenues respectively. In the present model, proposals will be defined so that if two projects are pursued simultaneously it is the joint outlay and joint return that registers and not the unadjusted sum of their cash flows.

There are several ways to achieve this proper indication of cash flows. When the programming model is discussed in the next chapter, the alternatives for netting simultaneous cash flows will be presented. For the present, however, two facts ought to be noted. First, these interactions among projects in terms of cash flows constitute a phenomenon apart from physical interrelationships among projects. It is not necessary that two projects bear some physical relationship to one another in order that they affect each other's cash outlays or returns

---

<sup>42</sup>See Marglin (1963, p. 56) for a statement on the importance of project interactions in the objective function.

or both. For instance, while the return from the investment program in the previous store-location example will mirror the results of the interaction between the two stores, the two locations are not physically mutually exclusive nor, even more certainly, is one physically contingent upon the other.

Second, these deterministic cash-flow interactions seem to have received little attention in the capital-budgeting literature. Some work by Hillier, Marglin, Reiter, Weingartner, and Williams and Nassar constitute exceptions to this statement.<sup>43</sup> Other treatments of capital budgeting under certainty do not take explicit account of the fact that the outlay entailed and the return received if two projects are pursued jointly may differ from the algebraic sum of their requisite outlays and earnings, respectively.<sup>44</sup> Insofar as these interactions among projects are significant in actual capital-budgeting situations, it is important that a model presuming to analyze such situations incorporate these interactions in its framework. The model in this study does take account of these cash-flow interrelationships. The exact way in which it accomplishes this will be seen in the next chapter.

Returning from this digression with the assurance that proposals will be defined so as to take account of cash-flow interactions among projects, the investment program's net return in a given period can be expressed simply in terms of the returns from the individual proposals. Let  $y_i$  be the number of units of the  $i$ th investment project (physical or financial) that is undertaken. That is,  $y_i$  indicates the extent to which the  $i$ th project is accepted -- a building is or is not constructed, four machines of type Z are purchased, two research programs are

---

<sup>43</sup>Hillier (1964, pp. 23, 70-73); Reiter (1963); Weingartner (1966, pp. 494-498); Williams and Nassar (1966, pp. 852-853).

<sup>44</sup>See, for example, Baumol and Quandt (1965, p. 320, equation [1]) and Weingartner (1963, passim). In the case of the more verbal discussions of capital budgeting it is difficult to cite specific references since the shortcoming is one of omission rather than commission. It is easier, therefore, to indicate where the mathematically oriented literature has erred since one can find specific equations that show the return from the program to be the algebraic sum of the individual projects' returns.

undertaken, K dollars are loaned in period  $t_0$ , and so on.<sup>45</sup> Assuming there are N potential investment projects -- including the real projects, artificial ones required to capture interaction effects, and market opportunities -- from among which the firm may choose, it follows that  $\sum_{i=1}^N a_{ti}y_i$  is the  $t$ th-period gross return from the investment program.

Let  $R_t^*$  denote the residual in the  $t$ th-period from the net earnings of the firm's initial resources after all investment outlays have been made. That is, in this model,  $R_t^*$  is the difference between the amount available for investment in period  $t$  from resources it controlled at  $t = 0$  and the gross outlays for the investment program in period  $t$ ,  $\sum_{i=1}^N c_{ti}y_i$ .<sup>46</sup> From our previous discussion, it follows that

$$(4.3) \quad W_t = \sum_{i=1}^N a_{ti}y_i + R_t^* .$$

The utility function to be maximized can consequently be written as

$$(4.4) \quad U = U \left( \sum_{i=1}^N a_{1i}y_i + R_1^*, \dots, \sum_{i=1}^N a_{Ti}y_i + R_T^* \right) .$$

To say that the objective of the capital program is maximizing the utility function in (4.4) does not exclude the possibility that the goal might be equivalent to maximizing discounted present value. Recall that  $V_t$  in (4.1) was the net cash inflow to the firm from its operations, including the investment program, in period  $t$ . But then it is seen that  $V_t = \sum_{i=1}^N a_{ti}y_i + R_t^* = W_t$  in the utility function in (4.4) when the definition of  $R_t^*$  is recalled. The discounted-present-value function in (4.1) is then a particular case of the general function in (4.4). Specifically,  $\sum_{t=1}^T D_t V_t$  is the case in which  $U$  is linear in the  $\sum_{i=1}^N a_{ti}y_i + R_t^*$  variables, with the coefficient of  $W_t$  being  $D_t$ . After arguing against the

<sup>45</sup>More will be said in the next chapter about the restrictions on the values the  $y_i$ -variables can assume.

<sup>46</sup>As one might expect, the  $R_t^*$ -values are themselves quite intimately related to the  $y_i$ -values. It is best, however, to postpone the precise statement of this relationship until the programming model is presented in detail.

discounted-present-value goal, have we defeated our own argument and allowed it to enter via the back door? The answer is no.

In the discussion of capital budgeting under risk, it will be seen that assumptions about rationality make certain very specific demands on the form of the utility function in (4.4). Even under certainty, though, some forms of the utility function seem implausible -- if not inappropriate -- for use in the capital-budgeting problem.

Among the recent advocates of the Fisherian approach to investment decisions, Hirshleifer does not present any specific functional form for the utility maximand, concerning himself only with indifference curves. Baumol and Quandt first state the utility function in a completely general form as that in (4.4). But then they consider a utility function additive in the utility of different periods and linear in money in each period. It takes the form

$$(4.5) \quad U = \sum_{t=1}^T u_t W_t ,$$

where  $u_t$  is the fixed utility of a dollar in the  $t$ th period and  $W_t$  is the amount available for withdrawal from the firm in that period.<sup>47</sup> They write that their "argument could have been made to appear more esoteric and profound by dealing with non-linear utility functions, bringing in concave programming, the Kuhn-Tucker theorem and all the associated paraphernalia. However, ... this would have added nothing to the argument."<sup>48</sup> For the argument Baumol and Quandt were making, the difference between (4.5) and a nonlinear utility function was one of form and not substance. But in terms of the more general question of what constitutes an appropriate objective function for capital budgeting, this difference is a substantive one.

Both the additive nature of the utilities of different time periods and the linearity of each individual period's utility of consumption income in the

---

<sup>47</sup>Baumol and Quandt (1965, p. 326).

<sup>48</sup>Ibid., footnote 2.



Baumol-Quandt function seem implausible.<sup>49</sup> The former assumption means that the utility of having \$k available for consumption in period t is independent of the purchasing power available in any other period t'. Interdependence of the utility of consumption incomes in different periods is thus ruled out by this property. But one can argue that the utility of an extra dollar of consumption income in the present period may well depend on how well the individual was provided for in the past and how well off he will be in the future. For example, if he has lived comfortably in the past, then any reduction in income may be more strongly felt than if his previous consumption incomes had been lower. Having grown accustomed to a particular way of living, the need to move to a lower level as income declines might actually be accentuated.<sup>50</sup> On the other hand, an individual's present consumption income might well be worth more to him if he knew that his future were better provided for rather than if he knew that future periods would find him in a less desirable position.

The thrust of the argument is that whether or not the dollars one has today interact directly with the dollars possessed tomorrow -- or in days thereafter -- in determining the utility of a consumption-alternatives stream will depend on the individual (or corporate ownership) in question. Such interactions are clearly possible and I would assert -- since neither theoretical proof nor empirical evidence can be adduced to settle the point either way -- more likely than not to exist. A multiperiod utility function that excludes the possibility of such interactions, namely, one additive in individual-period utilities, must at least be considered overly restrictive, if not implausible.

The assumption that utility is linear in money in each period has an even more unrealistic and undesirable result. It implies that the marginal rate of

---

<sup>49</sup>The additive property is equivalent to assuming the utility function is strongly separable with regard to consumption possibilities in different periods. For a further discussion of separability, see Goldman and Uzawa (1964).

<sup>50</sup>This is similar to the ratchet effect discussed in Duesenberry (1949).

substitution between dollars in any two periods is a constant. In particular, the function in (4.5) states that the marginal rate of substitution between dollars in period  $t$  and dollars in period  $s$ ,  $\left| \frac{\partial W_s}{\partial W_t} \right|$ , always equals  $\frac{u_t}{u_s}$  which is a constant. Owners are supposedly willing to trade dollars in one period for dollars in a second period at a constant rate. Ceteris-paribus indifference curves between consumption alternatives in different periods -- not just successive periods -- must be linear. This is highly implausible. The marginal rate of substitution between dollars in different periods should, one would think, depend on the actual consumption alternatives available in those periods. For example, the amount an individual would be willing to substitute in period  $t$  for \$100 in period  $s$  if his consumption income in period  $t$  were \$5,000 and in period  $s$  were \$9,000 would generally differ from the substitution he would willingly make if those consumption incomes were \$9,000 and \$5,000 respectively. The linearity of indifference curves that is implied by (4.5) seems to be a defect that, while not impeding the argument put forth by Baumol and Quandt, ought to be corrected. Such a remedy is to be found in introducing nonlinearities into the utility function.

Furthermore, it is only with a nonlinear utility function that Hirshleifer's conclusion about the dependence of the proper discount factor on the firm's optimal position comes into its own. Each marginal discount factor is the reciprocal of the absolute slope at the point of tangency of the relevant indifference curve and relevant productive- or market-opportunities locus between successive periods' dollars. The shapes of the indifference curves and the production-, borrowing-, and lending-opportunities loci between successive periods are known at the outset, but the marginal discount rates -- and hence the correct discount factors -- are only determined in the full solution to the problem. The linear utility function implies, in direct contrast, that these discount factors are known ex ante

since the indifference curves have constant slopes. Clearly, with a linear utility function some of the strength is sapped from Hirshleifer's conclusion.

In a recent paper, A. Manne raised another objection to the use of the linear utility-of-money function by Baumol and Quandt. He, however, abstains from taking issue with the additivity of the period-by-period utilities. His argument rests on the assumption of constant returns to scale for investment activities and the assumption that all projects are of the deterministic "point input-stream output" type. Each project has one period in which a positive input of cash is made followed by a stream of nonnegative cash throw-offs in successive periods.<sup>51</sup> For such a world Manne demonstrates that the Baumol-Quandt linear utility function leads to optimal solutions in which positive withdrawals during time  $t$  mean zero investments during that period and conversely positive investment in a period implies no withdrawal at that time.

This all-or-none theorem means that the B-Q [Baumol-Quandt] linear programming model is hard-pressed to explain a readily observable phenomenon: that well-managed enterprises typically invest and also pay out cash dividends [have withdrawals made] during the same fiscal year. Nonlinearity appears to be essential if a capital budgeting model is to explain the simultaneous occurrence of investment expenditures and cash dividends [withdrawals].<sup>52</sup>

Thus, for a model quite different from the one basic to the present work, but close to that of Baumol and Quandt, Manne concludes the need for a nonlinear maximand.

Before concluding this discussion of the objective function for capital budgeting under certainty, mention ought to be made of a significant recent article by A. C. Williams and J. I. Nassar.<sup>53</sup> They begin by presenting a set of properties that they believe are commonly associated with the notion that one

---

<sup>51</sup> Compare these assumptions with the nature of the investment projects discussed in this study, as described in Chapter 5.

<sup>52</sup> Manne (1966, p. 2).

<sup>53</sup> Williams and Nassar (1966).

investment is better than a second. These properties include: (i) "greed" or a preference for a cash flow A over B if A is at least as large in every period as the second cash flow B and greater than B in at least one period, (ii) "impatience" or a preference for A over B if the two cash flows are identical except that the amounts from A are each received one period sooner than the amounts from B, and (iii) "marginal consistency" or a preference for A over B if and only if the difference between them, A-B, considered as a separate entity is preferred to the cash flow of zero in each period. Two further axioms complete their set. The first is simply one of continuity: if cash flow A is preferred to B, then there exists a small enough but positive cash flow  $\epsilon$  such that if  $\epsilon$  is subtracted from A, A- $\epsilon$  is preferred to B. The second assumes there is no absolute reference for time: if cash flow A is preferred to cash flow B, then if the start of each is postponed one period, the new postponed cash flow A' is preferred to the postponed cash flow B'.

Williams and Nassar then show that the only method for ranking cash flows that is consistent with the axioms they present is according to discounted present value. In other words, they conclude that a preference ordering is consistent with their axioms if and only if it is the (linear) discounted-present-value objective function. Their result differs from the position taken here because their axiom of marginal consistency, property (iii) above, is equivalent to an assumption of constant marginal utility for income in any single period. This assumption, the preceding discussion has argued, is an implausible one from the present author's point of view.

This discussion of the goal of capital budgeting in a world of certainty but imperfect capital markets contains several conclusions that ought to be repeated as they are basic to what follows. First, it was argued that the objective function to be optimized ought to be a utility function. Second, the utility function in question would be taken to be management's perception of the owners'

utility function defined over consumption alternatives over the budget's T-period horizon. Third, the arguments of the utility function are the amounts available for withdrawal from the firm in the several periods, with these figures serving as proxy variables for the true increases in stockholders' consumption possibilities. Fourth, it was shown in equation (4.4) how, given proper definitions of proposals or projects, the objective function could be written in terms of the individual projects themselves. The discussion concluded with an argument that this utility function should be nonlinear. The nonlinearities might also occur between individual periods' utilities but definitely ought to occur within the individual period utilities.

#### 4.3. An Evaluation of Previously Suggested Rules for Budgeting Capital Under Certainty

The literature on capital budgeting provides a number of rules to aid decisionmaking units in their choice of optimal investment programs. Some of these suggestions remain in the realm of pure theory -- as our look at the way capital-budgeting decisions are actually made will show -- while others have been adopted by firms, government agencies, and the like. Since this study proposes at least a different way of viewing the capital-budgeting decision, it seems important to ask why the rules in the already voluminous literature, whether theoretical or applied, should be supplanted. It is to this question that we turn now.

Whether or not any particular rule is optimal can only be determined by examining its consistency with the decisionmaking unit's objective. The rules previously suggested do not lead the firm or division operating in the capital-market setting described in the previous chapter to a maximum value of the nonlinear utility function in (4.4). With the exception of the programming literature, and Hirshleifer's theoretical work, the literature's suggestions have one defect in common. They would have the firm construct capital programs through a one-by-one selection procedure. Each project would be individually accepted or rejected on

its own merits. The traditional search for investment rules is often even phrased in such a piecemeal way. For example, Dean and Smith write:

The central problem of capital budgeting is the choice of a criterion of capital productivity to measure the value of individual investment proposals.... The company's major capital-budgeting problem is to pick the best set of proposals that can be had with available funds, preferably by the use of an index of investment worth that is easy to apply and correct in its measurements.<sup>54</sup>

Selecting investment projects one by one presents no obstacle to reaching the optimum position so long as the proposals considered are completely independent of one another. This means more than the absence of physical interrelationships of the mutual-exclusion or contingency type. The presence of such physical interdependence certainly invalidates one-by-one approaches. For example, one project may appear worthwhile on the basis of the cash flows it requires and it generates, but it may, at the same time, preclude the acceptance of several other projects the sum of whose net cash flows is preferable to the single project's. Or, constructing a building may appear disadvantageous until one considers the profitability of a specific facility such a building would make possible. Considering the compound project consisting of the building and the specific facility may then reveal a profitable venture.

Interrelationships among projects can, however, arise in other ways as well. The interactions that can occur among projects in terms of cash outlays and returns have already been discussed. They constitute a very important type of interdependence that one-by-one approaches fail to take into account. Another set of interdependences arises because project indivisibilities in conjunction with the period-by-period budget constraints force the fate of one project to be dependent on the action the firm takes on other proposals. Suppose, for example, that in

---

<sup>54</sup>Dean and Smith (1955, p. 293). In a recent survey of the techniques used by business firms in reaching investment decisions, D. F. Istvan refers to "the measure of acceptability" as "a device for allowing management to determine how the advantages of a particular proposal compare with those of other proposals." Istvan (1961, p. 47). Italics are mine.

one of the periods the firm has \$100,000 to spend and that three of the indivisible projects it is considering entail outlays of \$60,000, \$30,000 and \$25,000, respectively, in that period. The proposal requiring the largest outlay may appear "worthwhile" in terms of the cash outflows it requires and cash inflows it generates over the horizon. But accepting it will force the firm to reject at least one of the smaller-outlay projects. If the combined returns on the smaller-outlay proposals are "better" than those of a combination of the big project and either of the smaller ones, the firm will be led astray by a one-by-one approach. This will be true even if the project involving the largest outlay provides the highest single "return" under whatever standard the firm may invoke.

Since the firm in the present model can borrow and lend it would, perhaps, have been better to phrase the difficulty just discussed in terms of an effective borrowing constraint as opposed to an effective budget constraint. The point remains the same in either case. The imperfection of the capital market in the form of divergent borrowing and lending rates does, however, create a problem all its own for the piecemeal approaches. If one is to evaluate proposals using the market rate of interest in the calculation -- as at least three of the rules to be discussed do -- the question arises as to which (if either) interest rate to use, the borrowing rate or the lending rate. The simultaneity of the discount-rate determination and the capital-budget allocation discussed earlier returns to haunt and confuse the firm that insists on deciding whether to accept or reject proposals one by one. It will be deciding whether or not to pursue a given project on the basis of a choice between  $r_L$  and  $r_B$ , but it will not know until after it has made its selection which, if either of them, is the relevant rate to use.

The degree of independence among projects that is required if judging projects individually is to be a valid procedure exists neither in the present model nor in the real world. Physical interrelationships among projects, interdependence in their returns and outlays, project indivisibilities,

internal-budget constraints, and imperfections in the capital market all stand in the way of a rational firm's use of a one-by-one approach.

But the rules and approaches that have been advocated elsewhere also fail the optimality test for a more fundamental reason. Perhaps this is stated too strongly. Rather, the rules that have been suggested will be optimal for use by the firm only under very special circumstances. Each of these rules implies a utility function for the firm to optimize. These approaches then fail -- if the argument presented here is correct -- because the utility functions they imply differ from the nonlinear one postulated in equation (4.4), as restricted by the discussion immediately preceding and following that equation. One of these methods could be correct only where the firm in question actually had a utility function of the form the rule implies. Since it has been argued that the firm should be optimizing the nonlinear utility function in (4.4), these other approaches must be judged nonoptimal unless the firm is, indeed, a special case.

#### 4.3.A. The Subjective Approach: Urgency and Necessity

The first major approach -- if one can really call it that -- is the subjective approach. It emphasizes the urgency or necessity of a proposal. As many urgent proposals are accepted as possible while postponable projects are postponed. The key question raised when this approach is employed is "To what extent can the project be postponed to a later time?" The measure has little place in a rational capital-budgeting scheme. Using it, the criteria for acceptance and rejection become more emotional than rational. The capital budget that emerges is more a reflection of the interplay of corporate personalities than anything else. This is not to say that some projects are not more urgent than others. But the urgency of a project will emerge in its time stream of returns, and it is precisely that stream the utility function in (4.4) takes into account. The "degree of necessity," on the other hand, is an immeasurable quantity that can tell us nothing.



It is difficult to say what, if any, utility function the subjective approach implies. Certainly, no rational account is taken of the stream of cash inflows and outflows from a particular project, let alone from a set of projects. The firm is taken to maximize some vague average degree of necessity or to minimize some average number of years of postponability in the program it selects. In any event, the implied optimization is definitely not equivalent to maximizing the utility function in (4.4) or, for that matter, any utility function that takes systematic account of the contributions of the several projects considered. For that reason the subjective approach is rejected.

#### 4.3.B. The Payback Period

A second criterion that has received considerable attention, both in theory (where attention does not mean advocacy here) and in practice, is the payback period. The payback period of a project is the number of years required for the earnings on the project to pay back the original capital outlay, with no allowance for depreciation or obsolescence of capital. A firm that employs the payback-period measure would aim at minimizing some average of the payback periods of the projects it selects subject to the constraints it faces. Alternatively, it would accept all projects its budgets allow it to that have a payback period less than or equal to some time period it sets.

The utility function implicit in this measure has some definite peculiarities. First, the utility of all net revenues from a particular project is positive for periods earlier than the payback limit and zero for all later periods. But this is unreasonable. Dollars received after the payback cutoff will have some utility, although probably less than those dollars received earlier. The payback criterion, however, ignores these delayed cash flows.

In addition, the payback criterion implies that the utility of a marginal dollar received in a particular period will vary from one project to another.

For example, suppose project A has a payback period of three years while B has a payback period of five years. A dollar received from project A in the fourth year would have zero utility while if that dollar were earned on project B it would have positive utility. This seems highly irrational for, as is well known, "a marginal dollar in period  $t$  is a marginal dollar in period  $t$  is ..." and so on.

Third, even for those periods in which the earnings from a particular project are valued, the payback measure's implicit utility function has a grave shortcoming. It assigns the same utility to a dollar no matter when it is received so long as it is received prior to the payback cut-off period. In short, the measure makes no allowance for time preference, ignoring as it does the time pattern of receipts except for the pre- and post-payback period demarcation. This would again appear to be highly irrational. There is an opportunity cost involved in postponing the receipt of money from one period to another, no matter how imperfect the capital market may be. Moreover, introspection and casual empirical observation would indicate that most people do display some degree of subjective time preference. Use of a rule that implies a utility function reflecting none of this preference, aside from a sharp zero-one break, can neither prove very helpful nor can it lead to an optimum position.

Another reason the payback period fails to lead to an optimum rests with its disregard of capital wastage. Since it uses gross earnings in its calculations, the rule fails to consider the obsolescence and depreciation of the asset. This is another element of its lack of foresight. The firm using the payback measure, in effect, does not care what accrues from a particular project beyond its payback period. Hence, the firm has no concern for the capital depreciation and degree of obsolescence of the asset. As for depreciation and the machine's "growing old" prior to the payback period, they are also of little relevance to the firm. All it cares about is how long it will take for the dollars flowing from the project to cover the initial outlay.

These properties of the payback measure would eliminate it as a means for generally achieving the optimization of the utility function postulated here. Aside from the payback approach's concentration on selecting projects on a one-by-one basis, the utility function it implies has very peculiar features. One can conceive of a situation in which emphasis might be put on rapid payoff, but it is still hard to envision a case in which a utility function of the type associated with the payback measure would be the overriding consideration. For example, if a firm is in an industry where technological change is occurring rapidly,<sup>55</sup> it may be important that investments pay their own way rapidly. Under such circumstances an emphasis on liquidity may be quite reasonable and signs point to the relevance of the payback-period measure. But even under such conditions surely overall profitability should remain the paramount consideration, a consideration which the payback criterion cannot take into account.

Note, however, that the utility-function approach advocated here can take account of both concerns -- the concern for general worth and the concern for liquidity and rapid return on investment. The returns in all periods would be considered since each period's increment to the stream of consumption alternatives makes its appearance in the function with nonzero weight. At the same time, an emphasis on liquidity would appear in the form of much larger utility weights on the earlier periods in the planning horizon. In short, the payback-period approach cannot serve as a general tool for optimizing the firm's position, and even in the case where its use is more justified it is not truly proper to substitute it for the utility-function approach so long as profitability is of any concern at all.

#### 4.3.C. The Terborgh or MAPI Formula

The Machinery and Allied Products Institute (MAPI) system developed by G. Terborgh typifies a third line of attack upon the capital-budgeting

<sup>55</sup>Recall that the discussion is still restricted to capital budgeting in a world of certainty. This forces upon us the untoward assumption that such technological change as occurs is completely predictable.

problem.<sup>56</sup> This approach embodies assumptions about the relevant aspects of the investment decision -- more specifically, the replacement decision in the case of the MAPI formula -- in a formula or chart. The chart then indicates the correct course of action when the necessary estimates are provided for the decision at hand. For example, after information concerning the amount of investment, the terminal salvage values of the equipment, tax rates, the "cost of capital," tax-depreciation patterns, and so on, are given, the Terborgh formula will tell the firm when and with what to replace a given piece of equipment. At the heart of Terborgh's method is the "adverse minimum," a measure of the average annual opportunity cost of any piece of equipment.

The total opportunity cost of a particular piece of machinery, be it presently owned or newly purchased, can be divided into two parts. The first part is the discounted present value of its acquisition cost. For presently owned equipment this is simply the present salvage value of the machinery -- the cost of keeping it. In the case of equipment to be acquired for the first time, it includes the discounted present value of original acquisition costs and future capital additions minus the discounted present value of the new machine's future salvage value. The second component is the "operating inferiority" of the proposal -- whether it is a new machine or presently owned equipment. This measures the equipment's handicap in comparison with the best machine available at any given moment. A proposal's "operating inferiority" thus equals the discounted present value of the stream of cost disadvantages that results if the project is pursued instead of using the newest and best equipment coming available each year.

As the assumed lifetime of a piece of equipment increases, both components of this opportunity cost increase. The discounted present value of the acquisition cost increases because the salvage value is declining and is being discounted more heavily. The operating-inferiority component increases (1) because technological

---

<sup>56</sup>Terborgh (1949); Machinery and Allied Products Institute (1950).

improvements will lower the first-year operating costs of new equipment and make their output better adapted to the changing pattern of future demand, and (2) because as the piece of equipment being considered depreciates with time, losing precision and requiring greater maintenance, the operating costs of the machine increase.

For any given lifetime of the machine, Terborgh expresses this total opportunity cost as an equivalent annual return or annuity. That is, if  $r$  is the discount rate (the "cost of capital") and the equipment's lifetime is from period 1 to period  $T_0$ , the total opportunity cost of the proposal is expressed in terms of the annual amount  $m$  where

$$(4.6) \quad m = (\text{Total Opportunity Cost}) \left[ \frac{r(1+r)^{T_0-1}}{(1+r)^{T_0}-1} \right]$$

The amount  $m$  is defined so that the total opportunity cost equals the discounted present value of  $m$  dollars received each period from period 1 to period  $T_0$ .

Terborgh then chooses the service life  $T_0$  that minimizes this average annual opportunity cost  $m$ . The resulting minimum value of the equivalent annual return, call it  $\bar{m}$ , is what is called the "adverse minimum." One then chooses the proposal, either the existing equipment or a piece of new equipment, with the lowest adverse minimum. It is to be replaced by a machine of the same type each  $\bar{T}_0$  periods where  $\bar{T}_0$  is the service life for which the adverse minimum is attained.

The two key assumptions underlying the application of the MAPI method are:

(1) that operating inferiority increases linearly over the life of the proposed investment, that is, due to deterioration and obsolescence, the cost of operating the equipment increases at a constant rate over time, and (2) that all future equipment of the same type as that presently considered will have the same adverse minimum as the piece of equipment presently considered. These assumptions are both overly restrictive and immediately lead one to question the

usefulness of the Terborgh formula. While the MAPI approach should be given credit for making the realistic assumption that expected returns from a proposal decline over its life, to assume that they decline at a constant rate is too stringent. There may be cases where such an assumption is justified but one cannot reasonably take it to be valid in the general case. Hence, while the Terborgh method does constitute an improvement over the subjective and payback-period approaches in taking some account of a time pattern of receipts, the account it does take remains severely inadequate.

This is compounded by the weakness embodied in its second key assumption. It is, clearly, not normally the case that the relative advantages of proposals will remain in the same ratio for all future years. Yet this is what the assumption of a constant adverse minimum implies. No matter how far out into the future one peers, if it is desirable to replace a presently owned piece of equipment by a particular new one now, it is desirable to continue using that new type of equipment at each replacement step in the future.

The Terborgh or MAPI formula is thus not a very useful tool for choosing the optimal investment program for several reasons. First, it is clearly a member of the one-by-one selection school. Second, it is primarily an instrument for effecting replacement decisions. It can be extended to more general investment problems but there are definite difficulties involved in doing this.<sup>57</sup> A major difficulty is that the assumptions objected to earlier in the case of replacement decisions become even more untenable in the realm of general investment decisions where replacing equipment is competing with a whole spectrum of capital activities. Third, the formula is inadequate in the presence of an imperfect capital market since it requires some measure of the "cost of money" in its calculation of the adverse minima of the different proposals. The rule here finds itself in the odd position of requiring at the outset information that will only be known after one has finally decided upon the composition of the budget.

---

<sup>57</sup>See, for example, Dean and Smith (1955).

The last defect leads back to the utility-function framework for evaluating previously suggested rules. The utility function implicit in Terborgh's rule is not easily unravelled. It is clear, though, that it is not the nonlinear function postulated in equation (4.4) and the accompanying discussion. The heavy reliance on the cost of capital in the discounting procedure, the simplistic assumption concerning the declining stream of returns, and the extremely heavy weighting of the first-period operating cost inferiority all serve to indicate that Terborgh's rule is not the same as the approach suggested here. This is not to say that there will never be cases in which the MAPI formula leads to the same result as the utility-function approach put forth here. But the cases in which the MAPI rule and the more general utility-function decision are in agreement will be rare. Only when the conditions of the case are exactly those assumed for reasons of simplification in the MAPI calculations will the two methods reach the same conclusion. When those simplifying assumptions are unrealistic, the MAPI-formula and the utility-function approaches will diverge with the latter course being the correct one.

In his discussion of the MAPI formula developed by Terborgh, D. F. Istvan writes that "Use of the charts and formula produces a rate of return roughly comparable to that obtained under the computation of the discounted cash flow."<sup>58</sup> One might then think that the utility function implicit in Terborgh's rule is possibly close to that implicit in what Istvan calls the discounted-cash-flow approach. If either of the methods that Istvan joins together under the latter heading were to provide an optimal rule for investment decisionmaking, the Terborgh formula would then appear to stand on somewhat firmer ground.

---

<sup>58</sup>Istvan (1961, p. 49). Note that Istvan uses the phrase "discounted cash flow" to cover both what is otherwise known as the "internal-rate-of-return method" (or "the investor's method" or "the scientific method" or "the marginal-efficiency-of-investment method") and the discounted-present-value rule. The majority of the literature reserves the term "discounted cash flow" for the former approach only. I shall follow the majority of the literature in this regard.

Unfortunately, neither the discounted-present-value rule nor the internal-rate-of-return rule is satisfactory. Neither of these approaches leads to the optimization of the utility function in (4.4). Let us see why.

#### 4.3.D. The Internal-Rate-Of-Return Rule

The discounted-cash-flow approach begins by calculating the internal rates of return of the candidate proposals. The internal rate of return or marginal efficiency of investment of a given proposal is that discount rate which makes the discounted present value of the stream of net cash flows from the proposal equal to zero. If the proposal is to buy machine A rather than machine B, the stream of net cash flows whose discounted present value one wants reduced to zero is the stream of differences in the flows of A and B. In contrast, if the proposal is to buy machine A or not to buy machine A then the internal rate of return of the proposal is the rate that discounts the stream of net flows into and out of A to zero.

The discounted-cash-flow method then instructs the decisionmaking unit to accept all proposals with internal rates of return in excess of the unit's cost of capital. If the firm or division or agency faces a budget constraint that prevents it from accepting all proposals with a marginal efficiency of investment greater than its cost of capital, the competing proposals would simply be ranked in order of their internal rates of return. The firm would then proceed down the list accepting proposals until the fixed sum was exhausted. Writers in the internal-rate-of-return school usually do not discuss a procedure involving their central tool when budget constraints exist in more than one period. In fact, it is only the proponents of other methods who have suggested how the internal rate of return might be used when a single budget limitation exists.<sup>59</sup>

---

<sup>59</sup>See Lorie and Savage (1955, pp. 63-66) for a discussion of the use of the internal rate of return under conditions of capital rationing. They also pay particular attention to the use of the marginal efficiency of investment in selecting from among sets of mutually exclusive proposals.



The method based on the internal rate of return does have advantages over the other rules or approaches that have been reviewed thus far. For one thing, it does force the decisionmaker to consider the entire economic lifetime of the proposal and to concentrate on the project's lifetime earnings and outlays. The method accurately reflects the timing of expenditures and receipts. But its advocates err when they say that "It weights the time pattern of the investment outlay and the cash earnings from that outlay in such a way as to reflect real and important differences in the value of near and distant cash flows."<sup>60</sup> The discounted-cash-flow approach does not accurately reflect the differences in the value of money in different periods.

In calculating the internal rate of return, each period's cash flow is on an equal footing with every other period's flow. The only discounting that takes place is the result of using the marginal-efficiency-of-investment figure that is determined in the process of applying the rule. But the value of money to the firm may differ from one period to the next, and the application of the discounted-cash-flow method completely misses these differences. Even if the capital market were perfect, this myopic characteristic could still appear.<sup>61</sup> Suppose the interest rate varied from one period to the next. This means that dollars in period  $t$  will differ in value from those in  $t+1$  by more than the simple application of the marginal discount factor based on the internal rate of return. Nevertheless, the internal-rate-of-return approach would treat equal sums in the two periods as being equally desirable except for the marginal-efficiency-of-investment discount.

The possibility of interperiod variation in interest rates also points to one of several operational problems involved in applying the internal-rate-of-return method. Suppose one does determine a real, unique internal rate of

---

<sup>60</sup>Dean (1954, p. 33).

<sup>61</sup>Hirshleifer (1958, p. 226) discusses this possibility.

return,  $\rho$ , for a proposal. The rule instructs the firm to compare  $\rho$  with its cost of capital. If the corporation must cope with an imperfect capital market the problems involved in comparing anything with the elusive "cost of capital" should be obvious from the argument presented earlier in the chapter. But grant the proponents of this approach the assumption of a perfect capital market. Applying their rule may still entail insurmountable difficulties. If the interest rate varies from one period to the next -- in a predictable fashion since certainty is still assumed -- how is one to combine the different rates into a single "cost of capital" figure with which to compare the internal rate of return? Nowhere is there any indication on the part of the discounted-cash-flow school as to how this calculation might be made.<sup>62</sup>

In noting the operational difficulty caused by interperiod variations in the interest rate, it was assumed that a real, unique marginal-efficiency-of-investment figure existed. But this need not be the case. To determine the internal rate of return, one seeks the roots of a polynomial. Specifically, one wants to determine  $\rho$  such that

$$(4.7) \quad g_{1i} + \frac{g_{2i}}{1+\rho} + \frac{g_{3i}}{(1+\rho)^2} + \dots + \frac{g_{Ti}}{(1+\rho)^{T-1}} = 0,$$

where  $g_{ti}$  is the net inflow from proposal  $i$  in period  $t$ . But the roots of this equation need not be real and even if there are no complex roots, the roots need not be equal. In short, there is the possibility, on the one hand, that no internal rate of return exists, and there is also the other chance that one obtains an overabundance of marginal efficiencies of investment.

The problem created by the nonexistence of any real internal rate of return is clear: there is nothing to compare with the cost of capital. If more than one rate exists, the dilemma is no easier: which one does the decisionmaker use in

---

<sup>62</sup>Ibid., p. 222. Hirshleifer (1958), the source of this point as well as of others mentioned here, indicates that Fisher himself never indicates any way to effect this computation.

evaluating the proposal? It may appear worthwhile to accept the proposal if one root of the equation in (4.7) is compared with the cost of capital while rejection of the proposal may be indicated if another root is compared with the terms on which the unit can acquire funds or can use them elsewhere.

These problems in determining an internal rate of return can arise when the present value of a proposal is not a monotonically decreasing function of the variable  $\rho$  in (4.7). Lorie and Savage indicate that such would be the case if a proposal were to have an initial cash outlay, subsequent net cash inflows, and then entail a net cash outflow again at the end of its life. They suggest such proposals occur especially often in the extractive industries where terminal costs may be significant.<sup>63</sup> It seems wise to guard against the view that projects that created difficulties for the internal-rate-of-return calculation are, in some sense, weird or strange. As Hirshleifer writes, and supports with examples, "perfectly respectable investment options may have no real internal rates" and "It is definitely not the case .... that all options for which the internal rate cannot be calculated are bad ones."<sup>64</sup> Moreover, Lorie and Savage and Hirshleifer indicate that such "recalcitrant" situations are likely to become even more numerous when project interdependences, as mutual-exclusion relationships, are taken into account.<sup>65</sup>

Another difficulty facing the user of the internal rate of return occurs in selecting from among mutually exclusive proposals and is quite damaging for the discounted-cash-flow approach. Suppose two proposals are being considered -- machines C and D. Assume, moreover, that each has a real, unique internal rate of return with machine C's greater than machine D's. To complete the picture, the two machines are to be used for the same purpose and hence the firm only wants

---

<sup>63</sup>Lorie and Savage (1955, pp. 63-64).

<sup>64</sup>Hirshleifer (1958, p. 225).

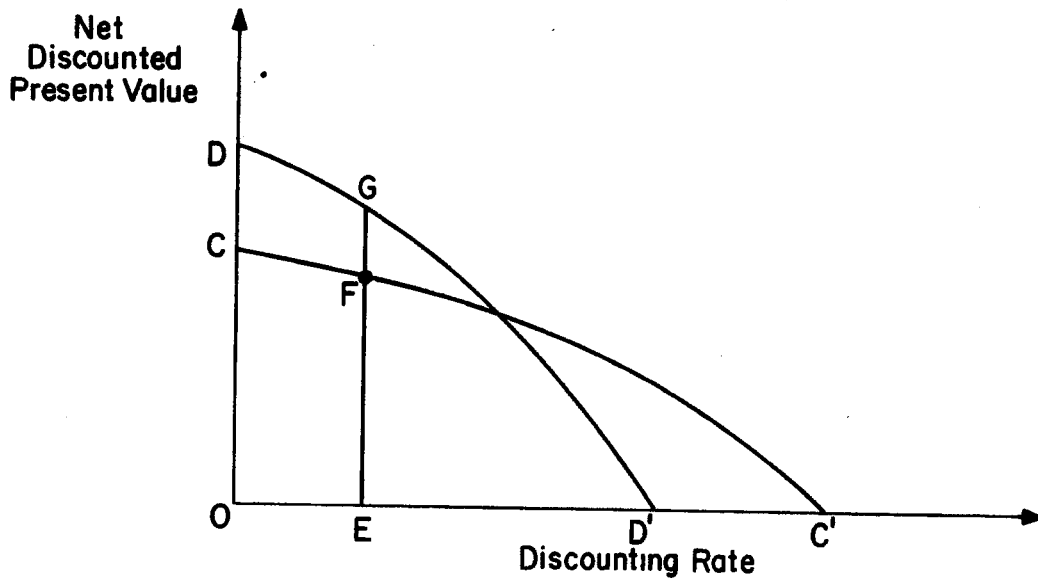
<sup>65</sup>Lorie and Savage (1955, p. 66). See also Hirshleifer (1958, pp. 217-219).

to purchase one -- if either -- of them; in short, they are mutually exclusive. It is not necessarily true that the firm should purchase machine C, even if both projects' rates of return exceed the cost of capital. Although C's present-value curve may cross the discounting-rate axis further to the right than D's curve cuts it, if the actual cost of capital -- assuming it is simply known -- is low enough, the discounted present value of machine D may exceed that of C. Figure 4.2 illustrates this point, where  $CC'$  is the curve relating the discounted present value of project C as a function of the unique discount rate and  $DD'$  presents the same information for project D.

Clearly, the internal rate of return of project C exceeds that of project D as  $OC' > OD'$ . But suppose the cost of capital to the firm is  $OE$ . Both projects' internal rates of return exceed the cost of capital so both would be accepted in the absence of a mutual-exclusion constraint. In the face of that constraint, the internal-rate-of-return rule would indicate that project C ought to be chosen. At a cost of capital equal to  $OE$ , however, the discounted present value of project D exceeds that of project C by amount  $FG$ . When the cost of capital is  $OE$ , the firm choosing project D could, if it wanted to, put itself in the position of adopting C by borrowing and lending and still have some of its wealth left for other uses! No rational firm would want to choose C under such conditions.<sup>66</sup>

The fundamental difficulty with the discounted-cash-flow procedure that lies behind all these shortcomings rests with an assumption implicit in the calculation of the internal rate of return. Specifically, it is being assumed that all intermediate cash flows, whether receipts or expenditures, should be compounded at

<sup>66</sup>For a similar discussion and a numerical example to satisfy the skeptical reader, see Hirshleifer (1958, p. 224). Also see Baumol (1965, pp. 443-446) for further discussion of this point. T. Marschak (1960, p. 136) indicates that it is this shortcoming of the internal rate of return that French planners find most disturbing, even above and beyond the possible nonexistence or nonuniqueness of the marginal-efficiency-of-investment figure.



**FIGURE 4.2**  
**Choosing Between Mutually Exclusive Alternatives**

the same interest rate  $\rho$  obtained in solving (4.7). But this is untenable. It will not generally be true that funds acquired for investment purposes will require interest payments of  $\rho$  percent each period. Nor is it necessarily true that the firm will be able to invest any cash inflows the project may give rise to in other proposals with a rate of return equal to  $\rho$ . The (internal or external) interest it will have to pay to finance the proposal will be calculated using the firm's "cost of capital" in the period in which the outlay occurs. The added return it can achieve from funds the project generates will depend on the other investment opportunities available to the firm in the period in which the inflow occurs. Only in rare instances will the mathematically obtained internal rate of return equal these two economically meaningful quantities.

Those writers who seek to preserve the discounted-cash-flow procedure suggest a remedy for this undesirable assumption. One of them, R. H. Baldwin, writes:

The funds would be at work during the interim period not at a rate similar to that of the proposed investment, but at the average rate at which general corporate funds are being invested -- at the over-all value of money to the company. An appraisal must be made by management as to what this value of money is now and what it is expected to be in the near future ... . It is at this rate that future cash flow must be discounted to reflect its present value in terms of the realities of the particular company's operations.<sup>67</sup>

Without taking the final step of saying it, these authors are actually calling for the adoption of a discounted-present-value rule. Cash flows in any period are to be evaluated on the basis of the opportunities available for their use, whether these opportunities are repayment of previous loans or new investments. For example, when Baldwin goes on to describe his new "real rate of return" measure, what he presents is, in essence, a discounted-present-value calculation which is then put on a rate-of-return basis. So it seems that those who sometimes come to praise but modify the internal-rate-of-return method actually conclude by burying it in favor of the discounted-present-value approach.

---

<sup>67</sup>Baldwin (1959, p. 99). Also see Solomon (1956, p. 79).

Given the criticism the approach has received, one might well ask if there is any case in which the discounted-cash-flow method is theoretically correct. One moment of partial, though not complete, success occurs for it when the firm has no budget constraint, and faces a perfect capital market, a continuously differentiable investment function, and a choice to be made among independent investment projects each with the same two-period horizon. In this case the rule will lead to the correct productive investment decision. That is, it will indicate correctly the physical investments the firm ought to make. Even in this case, though, the rule is silent on the market exchanges -- borrowing or lending -- that must occur if the firm is actually to reach an overall optimum. It might be said, then, that there is one case in which the internal-rate-of-return rule is theoretically correct as far as it goes, but that it does not go far enough even in this case.

The reason is not difficult to see.<sup>68</sup> As Fisher showed, the optimal investment decision requires, in part, that the marginal productive rate of return equal the interest rate between any two periods. But the marginal productive rate of return is simply the absolute value of the slope of the productive-opportunities locus, PP' or P''P''' in Figure 4.1, minus one. That is, the marginal productive rate of return equals  $-\frac{\Delta W_2}{\Delta W_1} - 1$  where  $\Delta W_t$  is the change in period t's consumption income, along the productive-opportunities locus. Hence optimization in the two-period case described requires that one set

$$(4.8) \quad -\frac{\Delta W_2}{\Delta W_1} - 1 = r ,$$

where  $r$  is the prevailing interest rate. Returning to equation (4.7), one finds that the internal rate of return for project  $i$  in the two-period case is

$$\rho = -\frac{g_{2i}}{g_{1i}} - 1 .$$

---

<sup>68</sup>For a fuller discussion of this point and of the success and failure of the discounted-cash-flow approach in different situations see Hirshleifer (1958).

And, in the present notation, the internal rate of return for total investment is

$$(4.9) \quad \rho = - \frac{\Delta W_2}{\Delta W_1} - 1 .$$

In short, in the two-period case with the capital-market and productive opportunities just described, the internal rate of return is simply the marginal productive rate of return. Considering this fact and the equilibrium condition expressed in (4.8), it is clear that all projects with internal rates of return exceeding the cost of capital,  $r$ , ought to be accepted.

This will lead the firm to the highest discounted-present-value position given its productive-opportunity locus. The firm is directed to the point of tangency between that locus and a market-opportunities line, since the slope of the latter is  $-(1+r)$ , and hence to the highest market line attainable. These market lines are given by equations of the form

$$(4.10) \quad W_1 + \frac{W_2}{1+r} = K ,$$

where  $K$  is the discounted present value along the particular market line. It can now be seen why, even in this case, the internal rate of return does not complete the task of choosing an optimal investment program. For only if the firm's indifference curves between dollars in the two periods are linear with slope equal to the discount factor has the rule led the firm to the highest indifference curve possible. But it was argued that the relevant utility function and indifference curves for the capital-budgeting problem will be nonlinear. Hence, even in this case, the internal-rate-of-return rule takes the firm only part of the way to an optimum. One must now by borrowing and lending move along this highest attainable line of discounted present value to arrive at the point of tangency between it and the highest indifference curve the firm can reach.

The discounted-cash-flow approach fails, in general, to select the optimal capital budget for the firm. As has been seen, the internal rate of return may



fail to exist in a particular case or, equally unfortunate, we may find ourselves with an embarrassment of riches in having too many rates of return. Problems in comparing the internal rate of return with the cost of capital may also arise on the cost-of-capital side where market imperfections or interest-rate variation through time may render us unable to choose a cost-of-capital figure. The difficulties inherent in the use of the discounted-cash-flow approach when rankings of projects are necessary were also discussed. More generally, the internal rate of return fails as a useful tool because it selects projects on a piecemeal basis as opposed to selecting the set of proposals that is best from the firm's point of view.

In the case where it takes the firm all the way to the optimum, the success of the rule based on the marginal efficiency of investment implies that a specific utility function exists. The firm's utility function must be linear with the respective discount factors as coefficients of the incomes in the different periods. Since such a function has already been discussed (or should I say inveighed against), there is no need to belabor its alleged shortcomings.

The utility function implicit in the discounted-cash-flow approach in general is even more peculiar. It is to maximize the absolute rate of growth of assets under the assumption that all intermediate returns from investments can be invested in further proposals with the same rate of growth. As Turvey writes, "Time preference and social discount are irrelevant here, since there is no choice between consumption at different points of time but merely a grim determination to plough back all surpluses and emerge at the end of the economic horizon with as large a stock of assets as possible."<sup>69</sup> Neither firms nor agencies in the real world nor the decisionmaking unit of the present model possess such a limited goal. The internal-rate-of-return approach must be rejected.

---

<sup>69</sup>Turvey (1963, p. 96).

#### 4.3.E. The Discounted-Present-Value Rule

What some, in fact, many, writers have suggested instead of the internal-rate-of-return approach is the discounted-present-value rule. This approach is also a piecemeal one in that it evaluates projects on a one-by-one basis. For that reason alone the rule would be deemed unsatisfactory, but its limitations go even further than its failure to look at groups of projects. The rule instructs the firm without budget constraints and existing in a perfect capital market to accept all projects with positive discounted present values and to reject all those with negative discounted present values. If budget limitations exist, the firm is to rank projects by discounted present value and use some iterative scheme based on these figures for accepting proposals.<sup>70</sup>

It remained for Weingartner, in his discussion of programming approaches, to make more explicit the nature and shortcomings of these previously suggested trial-and-error methods. Admittedly, in suggesting these iterative schemes the present-value advocates came close to considering sets of projects and moved away from choosing a capital budget in a one-by-one manner. But it was really only with the use of mathematical-programming methods that the selection of projects based on individual evaluations was fully replaced by the view that one should be selecting an optimal set of proposals.

The major difficulty in applying the discounted-present-value rule has already been discussed. Specifically, before a decisionmaking unit can use the rule, it must grasp hold of the elusive cost-of-capital figure to use in discounting future cash flows. The problems involved in this endeavor were discussed earlier. There is, in fact, only one situation in which the discounted-present-value rule can be said to be correct, and even then it is really only partially correct.

---

<sup>70</sup>See Lorie and Savage (1955); T. Marschak (1960, p. 136); and Renshaw (1957, p. 81).

This case occurs when the firm is not subject to any budget constraints and when it faces a perfect capital market. The discount rate to be used in putting future dollars on an equal footing with present dollars is then known ex ante. It is the single interest rate prevailing in the market.<sup>71</sup> If, given the absence of budget restrictions and the presence of a perfect capital market, there are no physical interrelationships among the firm's projects, the present-value rule will lead the firm to select that set of proposals that maximizes its discounted present value.

If two or more projects are mutually exclusive, the rule will still work if one uses the ranking modification. That is, the unit should rank all members of the mutually exclusive set on the basis of present value and choose that project with the highest discounted present value. It should be noted that this may lead to deviations from the "pure" discounted-present-value rule since some project may be rejected even if it has a positive discounted present value so long as it is in a mutually exclusive set with another project whose present worth is higher.

If, in addition, contingent-project relationships exist, the pure present-value rule requires even further modification if it is to lead to the maximum discounted present value for the firm. It may be that to maximize the present value of the firm a project with negative discounted present value should be accepted if it makes possible a project with a very large positive present value. This would be the case if the net contribution of the independent project-contingent project combination were sufficiently positive. This type of contingent relationship could be taken into account in applying the rule by considering the independent proposal-contingent proposal combination as a single compound project

<sup>71</sup>Clearly, if one wants to consider interperiod differences in interest rates, the previous sentence would have to be modified. The important figure would become the marginal discount rate for the specific future period. It would be known ex ante and would be the prevailing interest rate for that period. The crucial point is that, in contrast to the internal rate of return's problems, the presence of predictable variation in future interest rates in a perfect capital market creates no difficulty for the present-value rule.

that is mutually exclusive with the independent project alone. One could then simply invoke the method just described for dealing with mutually exclusive projects.

Hence, for a firm operating with access to a perfect capital market and subject to no budget restraints, the discounted-present-value rule, suitably modified when physical interdependence exists among projects, will lead to the maximum discounted present value for the firm. In contrast to the case with the internal-rate-of-return rule in such a setting, use of the discounted-present-value method will lead to this maximum value no matter how many periods are encompassed in the firm's horizon. But just as in the case of the discounted cash flow when it was correct, the discounted-present-value rule takes the firm only part of the way to the investment optimum. It indicates the correct set of productive investments for the firm to undertake but it says nothing about the borrowing and lending the firm should do in the different periods for which it is planning.

Once having attained the highest discounted-present-value hyperplane, the firm must still move along that hyperplane by borrowing and lending in different periods to reach the true optimum. The latter occurs when one of the firm's indifference surfaces is tangent to this highest discounted-present-value hyperplane.<sup>72</sup> Only if the firm's indifference surfaces were, in fact, coincident with these discounted-present-value hyperplanes would the discounted-present-value rule lead to the true optimum. But, as was argued earlier in the chapter, this will not normally be the case.

When one steps out of this perfect-capital-market environment, the rule is in even greater difficulty. The problem rests with the determination of the discount factor(s). But in an interesting paper, G. Pye has recently developed a set

---

<sup>72</sup>For a fuller explanation of this difficulty with the discounted-present-value rule in the two-period case when a perfect capital market exists, see Hirshleifer (1958).

of rules for accepting or rejecting different cash flows<sup>73</sup> (which may derive from a single project or a set of projects) when the firm faces an imperfect capital market similar to the one considered here. In Pye's model the firm can borrow and lend but the borrowing rate exceeds the lending rate. At the same time, in each period the firm has a fixed amount of income available from other sources. Pye's rules have a present-value interpretation relying as they do on his concepts of "generating present value" and "realizable present value." The former is the amount needed now to obtain a given cash flow over the horizon period while the latter is the amount that the given stream could be converted to at the present moment. (In a perfect capital market, the realizable present value and the generating present value of a given cash flow are equivalent. When the firm faces an imperfect capital market they are not, because the two market rates of interest are not equal.) As Pye emphasizes, "the rules that have been derived may neither accept nor reject a particular investment... When this is the case more information about preferences and other income will have to be provided before a decision can be reached."<sup>74</sup> In the theoretical framework assumed here, the utility function and the other income available are taken as fundamental although it is recognized that Pye's ingenious rules may well facilitate actual calculations.

The problem involved in attempting to apply the discounted-present-value rule in imperfect-capital-market situations in a more general and, in fact, entirely different way than Pye does is directly related to this underlying utility function to which Pye says we may have to return. Those who simply extend the present-worth approach from the perfect to the imperfect-capital-market situation --

---

<sup>73</sup>Pye (1966). Pye states that he seeks a rule to determine the acceptability of "projects." But his rules are, in fact, phrased in terms of preferred cash flows. The difference is more than a semantic one for a cash flow may be given rise to by a single project or by a set of projects undertaken in combination. Discussing cash flows rather than projects assumes that when projects interact with one another these interactions have already been taken into account in determining the cash flows. The approach taken in this study is to treat individual proposals as alternatives and to take account of interactions explicitly.

<sup>74</sup>Pye (1966, p. 50).

and Pye is not one of them -- usually invoke a discount rate that they say "expresses our valuation of futurity" or "properly reflects the investor's time value of money during [the] time period."<sup>75</sup> What they are, in fact, doing is postulating a constant marginal utility of money within each period and indifference surfaces that are hyperplanes. In short, they are assuming a linear multiperiod utility function of money for the decisionmaking unit.

This is most clearly seen from the work of Williams and Nassar referred to earlier. Recall that there a given set of axioms is shown to be necessary and sufficient for the discounted-present-value measure to reflect preferences accurately. One of these axioms, their Axiom III to be specific, is equivalent to assuming a constant marginal utility of income within each period.

This study has, however, argued that the objective function for the capital-budgeting problem is a nonlinear utility function. Since it implies a linear maximand, the discounted-present-value rule as it has been advocated for use in imperfect capital markets must thus be deemed nonoptimal. It is not being argued that if the correct discount rate(s) were somehow given in advance that the discounted-present-value rule could not then be used as a first step in approaching the optimum. Indeed it could be so employed, just as in the case of a perfect capital market where it takes the firm to the highest present-worth hyperplane. The subsequent step would then move the firm along that hyperplane to the highest possible indifference surface. Rather what is being said is that when its advocates have put it forth for use in imperfect capital markets they have proposed it as an end in itself. All time preferences have been thought -- or at least implied -- to be adequately captured in the discounting procedure. This is what is wrong because as an end in itself the discounted-present-value approach implies a linear multiperiod utility function. If the statements made earlier about the objective function are accepted, the discounted-present-value approach must then be rejected.

---

<sup>75</sup>Turvey (1963, p. 97) and Hillier (1964, p. 3), respectively.

#### 4.3.F. Previous Programming Models: The Need for Further Work

The last approach to capital budgeting that will be considered here is also the most recently developed one. It is the programming approach to capital-budgeting decisions. No simple rule or set of rules for accepting and rejecting projects emerges from the literature presenting this method. Instead those who discuss it demonstrate how the tools of mathematical programming can be brought to bear on the capital-budgeting problem.<sup>76</sup> Their concern is with establishing a programming framework for making capital-budgeting decisions and showing how the various algorithms available can then be employed to solve the resulting programs.

In the sense that a solution procedure is given, one might say the programming advocates present a method for making the capital-budgeting decision. But the methods that emerge from their writings are very different from the simple rules -- actually deceptively simple rules -- proposed previously. The programming methods are more general and more complex, and they do not suffer from many of the shortcomings of those other rules. Indeed, Weingartner's volume uses the programming approach as a setting in which to evaluate the methods previously suggested in the literature. In particular, using the programming model of capital budgeting, Weingartner is able to demonstrate some of the difficulties that arise in using the internal-rate-of-return and discounted-present-value rules that were discussed earlier.<sup>77</sup>

The major advance that came with the programming literature in capital budgeting was the presentation of a method concerned with evaluating investment programs as single integral entities. In this literature the emphasis shifted

<sup>76</sup>The most complete work setting forth the variety of ways in which mathematical programming tools can be used is Weingartner (1963). Others who have participated in the discussion of the programming approach to capital budgeting under certainty are Baumol (1965, Chapter 19); Baumol and Quandt (1965); Charnes, Cooper, and Miller (1959); Manne (1966); Marglin (1963); Reiter (1963); and Weingartner (1966). As shall be seen in a later chapter, other writers have indicated how programming methods would be useful when risk is introduced.

<sup>77</sup>Weingartner (1963, Chapters 3, 4, 8 and 9).

from accepting or rejecting individual proposals to selecting the overall optimum portfolio of capital-investment projects. It is, of course, one thing to describe a new framework for viewing a problem, quite another to present a means for solving the problem within that new framework. In this regard the case of those suggesting a programming outlook was strengthened by the existence of programming algorithms that could help effect this choice of a budget.

Framing the budgeting decision as an attempt to optimize some criterion function subject to a set of restrictions enabled the programming writers to recognize explicitly the constraints subject to which the firm must make its decision. Budget constraints, project indivisibility, physical interdependence among projects (mutual-exclusion or contingency relationships), capital-market imperfections of all types, and personnel constraints could all be explicitly and formally incorporated into the programming model. This contrasts sharply with the more or less ad hoc procedures previously suggested rules had used to cope with these realities of many capital-investment situations. By considering all these constraints simultaneously and by taking the maximand to be a function of all projects simultaneously, the programming approach pointed the way to a correct solution procedure for the capital-budgeting problem.

Since much has already been written on mathematical programming methods applied to capital budgeting under certainty, one might ask why the question should be pursued any further here. The answer is that while previous advocates of the programming approach have pointed the way, they -- for the most part -- have not completed the task they set for themselves. In a very basic sense, their writings fall short of delineating a theoretically correct framework for attacking the capital-budgeting problem.

The difficulty rests, in part, with their specification of the constraint set for the particular problems they discuss, but even more important are difficulties with the objective functions in their models. The problem common to their



treatment of constraints and objective function is the lack of attention paid to interactions among projects in terms of cash outlays and revenues. With the exception of Reiter's article and Weingartner's elaboration of Reiter's method,<sup>78</sup> the programming literature has nothing to offer in the way of advice for dealing with such deterministic cash-flow interrelationships. This stands in contrast to the explicit suggestions that have been made for taking account of physical interdependences within the programming framework. In the next chapter something further (hopefully, something useful) will be said about how to incorporate such deterministic cash-flow interactions into programming models.

Apart from this difficulty, the constraint sets used in previous discussions have suited their purpose very well. Indeed, the constraint set to be used in the formulation here is quite similar to the sets of restrictions found in the previous programming literature on capital budgeting. The criterion functions that have been used, however, do not provide an example after which further work on the programming approach ought necessarily to model itself.<sup>79</sup>

In the first article to demonstrate how programming methods could be used in budgeting capital, Charnes, Cooper, and Miller considered the buying, selling, and storage problem of a warehouse manager who has to operate within certain liquidity constraints. They used an inadequate criterion function as they maximized undiscounted cumulative profits.<sup>80</sup> In their objective function, a dollar received in year  $s$  was just as valuable as a dollar received in year  $t$ , where  $s$  and  $t$  were any two years. In short, the corporation (or its owners or its management) had no time preference at all. This would be a rare corporation indeed, and not one that I should think we want to set at the center of a theoretical discussion

---

<sup>78</sup>Reiter (1963) and Weingartner (1966, pp. 494-498).

<sup>79</sup>The exceptions to this statement are, as shall be seen presently, Baumol and Quandt (1965) -- given a sympathetic reading of their footnote 2 on page 326 -- and Manne (1966).

<sup>80</sup>Charnes, Cooper, and Miller (1959, p. 231).

of capital budgeting. This is not to deny the important contribution Charnes, Cooper, and Miller did make. To show how programming techniques could be used was certainly an important step forward, both substantively and as a suggestive force for other writers.

Unfortunately, the objective functions that have been used by most of those other writers still leave something to be desired. The usual course is to set out maximizing the discounted present value of the capital-investment program. This occurs in discussions of both corporate capital budgeting and public-investment programs.<sup>81</sup> Since the present study's case against the discounted-present-value maximand in an imperfect capital market has already been given, there is no point to repeating it here. Let us simply say that if the argument presented is accepted, previous programming models that have maximized discounted present value must be found wanting.

Another goal that has been put forth in the programming literature is to be found in Weingartner's horizon model.<sup>82</sup> In this formulation, the firm aims "to maximize the net value of assets, financial and physical, as of the horizon, where the former are expressed in terms of the funds available for 'lending' at that time, and the latter are represented by the discounted [to the horizon] streams of net revenues past the horizon."<sup>83</sup> It is easy to show that when this goal is assumed for a firm facing a perfect capital market, the result is equivalent to

---

<sup>81</sup>For examples of the former, see Weingartner's initial presentation of the Lorie-Savage problem in Weingartner (1963, Chapters 3 and 4) and Baumol's expositional example in Baumol (1965, pp. 448-453). Note, however, that Baumol himself raises questions about this objective function both in footnote 11 on page 450 of the previous citation and in Baumol and Quandt (1965). Marglin (1963) provides an example of the use of a discounted-present-value objective function in the social-investment setting.

<sup>82</sup>Weingartner (1963, Chapters 8 and 9).

<sup>83</sup>Ibid., p. 141.

maximizing the discounted present value of the firm.<sup>84</sup> The discounted present value is calculated using the prevailing market interest rate as the discount rate.

Its equivalence with discounted-present-value maximization might leave one uncomfortable with this horizon-value formulation since explicit maximization of utility appears to be ignored. But Weingartner is careful to indicate that since the firm faces a perfect capital market, the productive-investment decision can be separated from the financing (borrowing and lending) problem. After the former decision is taken, and the highest present-value hyperplane attained, the latter problem can be solved so as to reach the highest possible utility level. In short, Weingartner clearly recognizes and explicitly notes the second step so emphasized by Hirshleifer as necessary for reaching the optimal investment decision. His horizon-value model is then intended as the penultimate rather than the final step in optimization for a firm operating in a perfect capital market.

When he turns to a discussion of imperfect capital markets -- the primary concern of the present study -- Weingartner can offer much less justification for

<sup>84</sup> Weingartner uses duality theory, Ibid., pp. 143-146, in what I think is an incorrect way to prove this fact. Actually no such heavy apparatus is required. All we need do is note that since there is a perfect capital market, the firm maximizing discounted present value will never let funds in any period rest idle. Whether they are generated by the investment program itself or elsewhere in the firm, such funds will be loaned with the market rate of interest as the return. Hence, no money will be removed from the firm prior to the horizon period. The firm's net revenues in the first T-1 periods will all be zero. The discounted present value of the firm in the first T periods will then equal a constant,  $(1+r)^{-T}$ , times the amount available for lending in period T. That is, it will equal the compound discount factor from period 1 to period T,  $(1+r)^{-T}$ , times the net value of the firm's financial assets at the horizon. The discounted present value of the firm in the post-horizon period will simply be the same constant times the post-horizon returns of the investment program discounted to period T, or the compound discount factor times the net value of the firm's physical assets at the horizon. Thus, the total discounted present value of the firm equals a constant times the net value of its financial and physical assets valued at the horizon. Maximizing horizon value is thus equivalent to maximizing discounted present value when a perfect capital market exists.

this horizon-value model. He recognizes this problem explicitly and writes the following.

There is, of course, one further difficulty with any optimization model for a situation involving imperfect capital markets. This is the question of what is the appropriate function to be maximized .....

Without attempting to solve the problem of the appropriate criterion we shall here simply assume that the criterion is given from outside, and, for convenience, as well as to preserve continuity with our earlier formulation, we shall continue to maximize the terminal value.<sup>85</sup>

The presence of an imperfect capital market makes untenable -- as was argued earlier in this chapter -- the separation of the investment decision from the consumption decision. One cannot simply maximize discounted present value -- if one can even define it for the imperfect capital market in question -- and then optimize with respect to the consumption decision. And even if one could define discounted present value in this context, maximizing terminal or horizon value need no longer be equivalent to maximizing this discounted present value, as Weingartner notes. At the close of the footnote just cited he also states that formulations alternative to his horizon model could be used. It seems that in the context of an imperfect capital market, an alternative formulation should be used. In particular, further attention ought to be paid to the question of the appropriate criterion function and to a model based on the results of that inquiry.

The previous discussion in this chapter led to a model in which the non-linear utility function in (4.4) is the maximand. In this context, Weingartner's horizon model for imperfect capital markets seems overly restrictive for, essentially, it proposes a utility function attaching zero weight to all returns but those from period  $T+1$  onward. Only the post-horizon returns from the pre-horizon projects and the returns from post-horizon investments -- financial or physical -- are taken to be important. Hence, while Weingartner's model does demonstrate

---

<sup>85</sup>Weingartner (1963, pp. 158-159, footnote 1).

interesting and important features of capital-budgeting problems, there remains a serious question about the appropriateness of the objective function he uses.

This brief view of the programming literature on capital budgeting closes with recognition of the programming precursors of the present model. Although Baumol and Quandt use an additive linear utility function to present their argument, they clearly recognize the possible nonlinearity of the relevant utility function.<sup>86</sup> A sympathetic reading of the footnote just cited thus puts the Baumol-Quandt article in a more favorable light than the other programming literature with respect to the position taken here on an appropriate objective function. Certainly, at least their emphasis on a utility formulation was an important step away from the mainstream of the other programming literature.

But the author coming closest to the objective function presented here was A. Manne.<sup>87</sup> As will be recalled from the earlier discussion, Manne proposed a nonlinear utility function to avoid a rather unrealistic implication of the Baumol-Quandt model. Specifically, Manne postulates a nonlinear utility function,  $U(W_1, W_2, \dots, W_T)$ , that is differentiable and concave with each period's withdrawal having positive marginal utility. Although, he comes to it via a different path, Manne still reaches the same objective-function formulation as the present study does.

But the nonlinear function he introduces is only required to eliminate the "all-or-none" problem he finds in a specific linear programming model of capital budgeting when projects are of a specific type. One is left to wonder, then, how strongly (if at all) Manne would advocate a nonlinear utility function as the capital-budgeting objective in a model as the one in this study when not all projects are of that specific type. The position of the present study seems to be taken at a more fundamental level, arguing as I have for a nonlinear

---

<sup>86</sup>Baumol and Quandt (1965, p. 326, footnote 2).

<sup>87</sup>Manne (1966).

utility function as the goal regardless of the nature of the projects considered. Thus, while both Manne and I are led to the same objective function, we arrive there for different reasons. In a formal sense, though perhaps not in a substantive one, Manne's objective-function formulation comes closest to the model presented here.

#### 4.4. Some Concluding Remarks

The first part of the chapter considered the goal of the firm's capital-investment program. It was argued that when the corporation faces the imperfect capital market described in Chapter 3, the appropriate objective function for the capital-budgeting problem is the nonlinear utility function presented in equation (4.4),

$$(4.4) \quad \text{Maximize} \quad U = U\left( \sum_{i=1}^N a_{1i} y_i + R_1^*, \dots, \sum_{i=1}^N a_{Ti} y_i + R_T^* \right) .$$

The utility function is management's perception of how the owners value the increments to their consumption alternatives provided by the firm during the T periods. It was argued that the function should be nonlinear in each of its arguments,

$$\sum_{i=1}^N a_{ti} y_i + R_t^* \quad \text{for } t=1, \dots, T.$$

The degree of success--or failure --that previously suggested approaches could be expected to have in achieving this goal was then examined. Although the cards may have seemed pre-stacked in its favor, a programming approach with the nonlinear objective function given in (4.4) emerged as the best of the alternatives examined. Each of the other methods considered was seen to be unable to lead the firm of the present model to its optimal capital budget.

In the next chapter, we construct the nonlinear programming model that will direct the corporation operating in the financial and physical environment set out earlier to its optimum investment position.

## CHAPTER 5

### THE CAPITAL-BUDGETING PROBLEM UNDER CERTAINTY:

#### A NEW PROGRAMMING APPROACH

##### 5.1. Introduction

Having examined the previously suggested approaches to capital budgeting under certainty in the light of the utility formulation advocated here, it is time to present the nonlinear programming model incorporating the utility objective function. Before doing so, it is important to recognize explicitly the completely normative nature of this model. This normative role is not peculiar to the present work, although it is even more true of this study than of previous theoretical suggestions. In the preface to their book, which was explicitly intended to express the economist's approach to capital budgeting in the language of businessmen, Bierman and Smidt write, "Businessmen have tended to make capital budgeting decisions using their intuition, rules of thumb, or investment criteria with faulty theoretical foundations and thus are apt to give incorrect answers in a large percentage of the decisions."<sup>1</sup> They note the small impact the economists' theoretical and technical writings have had on the business manager.

The present chapter begins by reviewing some previously reported findings on how corporations actually plan their investment programs. This review will serve to place the programming model to be presented in its proper normative perspective. Following the discussion of these empirical findings, the nonlinear programming model for capital budgeting in the certainty environment that has been posited is developed. While the model presented here closely resembles previous

---

<sup>1</sup>Bierman and Smidt (1966, p. v). The quotation is from the Preface of their First Edition. They reprint that preface in their Second Edition (1966) and give no indication that the situation has improved significantly since the First Edition appeared in 1960.

programming approaches to the capital-budgeting decision -- and is, in fact, identical to these previous constructs in some respects -- several characteristics distinguish it from its predecessors. First, there is the appearance of a nonlinear utility function as the maximand that has been discussed at length in the previous chapter. Second, there is the specific period-analytic view of the present model. Next among these distinguishing features is the way the present model incorporates cash-flow interactions among projects. Finally, the precise capital-market opportunities and the disbursement option the firm is assumed to possess set this model apart from others that have appeared in the literature. All of these features of the present model will be given special attention in this chapter.

#### 5.2. The Normative Nature of the Present Model: A Review of Some Findings on the State of the Art of Capital Budgeting

Surveys of the capital-budgeting procedures of corporations lend support to the earlier quotation from Bierman and Smidt's work about the inadequacy of the techniques businessmen use in making investment decisions. For example, Istvan, concludes from his 1958 interview survey of forty-eight large firms selected from the Fortune Directory that the "majority of the nation's largest business firms fail to employ a theoretically sound approach to the expenditure of funds for capital investment."<sup>2</sup> He concludes that this failure is not the result of excessive costs of implementing what he considers theoretically correct (either discounted-cash-flow or discounted-present-value) methods. Rather it is due to the businessmen's "inability to understand the concepts and/or to promulgate an understanding throughout the various strata of the firm."<sup>3</sup> Curiously enough, Istvan found that in most of the firms he

---

<sup>2</sup>Istvan (1961, p. 45). The firms in his survey accounted for almost 25% of the total \$33 billion expenditure on plant and equipment reported by the Department of Commerce in 1959.

<sup>3</sup>Ibid., p. 51.



interviewed, the dollar advantages of the proposals and the amounts of the investments were being estimated and computed properly. The originators of proposals generally were also very careful to consider the most logical possibilities for achieving the desired purpose.

The shortcomings of corporate capital-budgeting procedures would have to be ascribed, he concluded, to the measures of acceptability used by firms. Recalling the earlier discussion of these procedures, consider that four of his forty-eight firms relied on an urgency-postponability criterion, two used the Terborgh MAPI formula, while thirty-seven firms used the payback period or its reciprocal, the simple rate of return (relating annual dollar advantage to total investment)! Only five of the forty-eight firms used what Istvan called the "time-adjusted rate of return," under which umbrella he includes both the discounted-present-value and internal-rate-of-return procedures.

Of course, Istvan's study is almost ten years old, although it appeared in the literature only six years ago. One would hope that business acumen would have sharpened since that time, in part as a result of the additional resources businessmen and economists have devoted to the investment problem. It is to be hoped that when the returns are in from the numerous studies presently being conducted, business will achieve a higher score.

A study more recent than Istvan's does not, however, provide support for such optimism.<sup>4</sup> In this survey, carried on in 1962 and reported in 1965, questionnaires were employed to study the capital-budgeting decisions of firms in the chemical industry, an industry considered to be relatively progressive. In addition, the survey procedure weighted the sample toward the largest firms in the industry. Even with this element of nonrandomness, Pullara and Walker found the corporations in the chemical industry slow to adopt the approaches the technical literature on capital budgeting advocated. For example, of

---

<sup>4</sup>Pullara and Walker (1965).

forty-four specific replies to the inquiry about the primary measure used in investment evaluation, twenty firms indicated that they relied on subjective judgment, nine indicated the payback period, while eight said they used the simple rate of return (unity divided by the payback period). Only seven of the forty-four made use of the internal-rate-of-return or investor's method. Apparently, none of the firms even progressed to the stage of using discounted present value as their primary measure.<sup>5</sup> The non-specific replies (sixty-three firms of the one hundred fifty to whom questionnaires were sent returned the forms), which were not tallied, supported the importance of subjective judgment in making the investment evaluation. And this phenomenon was not peculiar to the smaller firms.

Pullara and Walker do note that small- and medium-sized firms rely heavily on subjective judgment with larger firms using the simple rate of return, the investor's method, and the payback period as much as subjective judgment. In fact, they make the "rough estimate" that firms using the internal-rate-of-return approach (despite its shortcomings, it is better than subjective judgment alone) as a primary or secondary instrument in investment planning account for about 35% of the chemical industry's total investment. Yet they conclude that "Judgment, or judgment plus payback, is the overwhelmingly popular method among small- and medium-sized firms and accounts for roughly half of even the large firms."<sup>6</sup>

A final interesting observation from their survey was the lack of apparent correlation between the methods of evaluation a firm was using and its reasons for preferring these methods. Many of the firms did not even give reasons for using the methods they were employing. On the other hand, while firms employing the internal-rate-of-return approach all explained their preferences, they each gave different reasons for its use.

---

<sup>5</sup>Ibid., p. 404.

<sup>6</sup>Ibid., p. 406.

The common shortcomings of most of the methods that are actually used in making capital-budgeting decisions are their failure to take adequate account of the economic value of time and their failure to focus on choosing an optimal program of investments rather than just selecting projects one by one. The earlier discussion of various approaches that have received attention in the business and technical literature indicated that some methods were particularly poor on both counts. Yet these -- subjective judgment and the payback period (or its reciprocal)-- are precisely the methods being used most widely.

The most sophisticated tool corporations seem to use is based on the internal rate of return, although, according to Istvan, some use discounted present value as the main instrument. Although the former method is better than subjective judgment or the payback approach, it still has the sometimes severe shortcomings noted in the previous chapter's more theoretical discussion. The discounted-cash-flow approach falls short of optimality due to both its piecemeal approach and its lack of appropriate time perspective.

The method based on discounted present value is the best of all approaches being used by firms, but again it has its difficulties as was indicated earlier. First, choosing projects on a one-by-one basis remains at the center of this approach. Second, no clear statement is given as to how firms using this method arrive at their discounting factor. The elusiveness of the appropriate discount rate in a world of imperfect capital markets, as the ones these firms operate in, may explain their reticence. T. Marschak's report on French investment planning is interesting to note in this regard. He writes that in these industries "Ranking criteria other than present value are rejected everywhere."<sup>7</sup> But the problem of determining the "correct" discount rate is not really solved. Instead, the planners try to set some reasonable range of rates and perform something of a sensitivity analysis for "It may be possible ... to state that

---

<sup>7</sup>T. Marschak (1960, p. 136).

a certain project has positive present value per franc or has greater present value per franc than another project only for 'absurd' rates that are clearly higher or lower than the socially correct one."<sup>8</sup>

Given that the discount-rate figure is difficult to grasp when an imperfect capital market exists, one must recall the undesirable features of a discounted-present-value maximand if it is assumed to be capturing corporate or social time preference. Use of a discounted-present-value approach by firms and government agencies facing such markets must therefore be rated unsatisfactory because of the method's poor account of time perspective as well as its project-by-project focus.

The present study's discussion of capital budgeting in a world of certainty but imperfect capital markets rejects the project-by-project approach to capital-budgeting decisions and attempts to give time preference the important position it merits. The study emphasizes improvements upon the way in which investment decisions are now made with regard to two of the more glaring shortcomings of present procedures. It should be clear that there is no illusion that the approach suggested here is fully operational. The hope is, instead, that by setting the problem in this way, questions of capital budgeting may at least be asked in a different, presumably better, way and the path paved to better answers. With this normative goal in mind, let us now turn to a presentation of the capital-budgeting programming model.

### 5.3. The Programming Model's Period-Analytic View and Its Treatment of Net Returns from Investments

The decisionmaking unit of the model is an existing corporate enterprise planning an investment program for the next  $T$  periods. The resources the firm controls before the investment program starts (that is, at  $t = 0$ ) are expected to generate returns over the firm's investment-planning horizon. They are expected to keep the corporation an ongoing institution no matter what investment plans

<sup>8</sup>Ibid., p. 135.

the firm makes or how the chosen projects fare. Specifically, in each of the first  $T-1$  periods the firm anticipates a cash throw-off of  $C_t$  dollars,  $t=1, \dots, T-1$ , from firm operations apart from revenues from the investment program. In addition, the enterprise has at hand  $C_0$  dollars at the start of its investment-planning horizon. This money represents internal funds that have been generated prior to the start of the investment plan and that are available for use in the capital program in period 1.

In the present model a period-analytic view of the firm is taken. All outlays in a period -- for either the unit's ongoing operations or its investment program -- are assumed to be made at the start of the period. All cash inflows are supposed to occur at the end of the period. In particular, the anticipated cash throw-off in period  $t$  deriving from the firm's original resources accrues to the firm only at the end of the  $t$ th period. Hence, in the  $t$ th period the enterprise has available new internally generated funds in the amount  $C_{t-1}$  for use in its investment program. The corporation's capital-budgeting problem is to allocate the sequence of funds,  $C_0, C_1, C_2, \dots, C_{T-1}$ , among the available productive and financial investment proposals.

Some explanation and justification of this period approach is in order. In the case of the certainty model there is no reason why such an approach must be taken. But when we turn to the problem of capital budgeting under risk there is good cause for using a model with such a period-analytic view of the firm. In the interests of continuity of presentation the problem of capital budgeting under certainty will therefore also be presented within this period-analytic framework.

The difficulty that would arise in the risk model if such a view of the firm were not taken would occur in the budget constraints of the problem. If a period-analytic framework, with disbursements made (only) at the start of the period and receipts occurring (only) at the end of the period, is not adopted,

the constraint matrix in the programming model must contain stochastic elements. That is, writing the general programming problem with nonlinear objective function and linear constraints (ignoring for now any further nonlinear constraints on the variables that will actually appear in the model to be presented) as

$$(5.1) \quad \text{Maximize } f(x) \quad \text{Subject to } Ax \leq b, x \geq 0,$$

some of the elements of the matrix A would have to be random variables.<sup>9</sup>

While the literature on stochastic programming has made considerable progress in coping with the case where the requirements vector b is a vector of random variables with known distributions,<sup>10</sup> the efforts devoted to the case of stochastic technological coefficients (A-matrix) have borne little fruit. For example, in his chapter surveying the area of stochastic programming, Hadley devotes some attention to the programming problem when random variables appear in the A-matrix. After formulating the problem, he concludes, "Given the current state of nonlinear programming there does not seem to be any general method for obtaining the global maximum ..., although sometimes a local maximum can be found."<sup>11</sup>

Advocates of the chance-constrained programming approach to programming problems containing random variables would take issue with this negative evaluation.<sup>12</sup> They would say that chance-constrained programming has successfully coped with the problem of technological coefficients that are random variables. It is true that within the framework of chance-constrained programming the question of stochastic technological coefficients is treated the same way as a

---

<sup>9</sup>In (5.1), x is a column vector, b is a column vector, and A is a matrix. The objective function f(x) is a real-valued function of the vector of variables, x.

<sup>10</sup>See, for example, Dantzig and Madansky (1961); Elmaghraby (1959, 1960); Madansky (1962, 1963); and Williams (1965, 1966).

<sup>11</sup>Hadley (1964, p. 170).

<sup>12</sup>For a description of chance-constrained programming and its applications see Charnes and Cooper (1959, 1962, 1963); Thompson, Cooper, and Charnes (1963); and Naslund (1964). Two authors have suggested the use of chance-constrained programming in models of capital budgeting under risk: see Hillier (1964, Chapter 6) and Naslund (1964, Chapter 6; and 1966).

stochastic requirements vector (b in (5.1)). Nevertheless, a serious problem remains with the value of the chance-constrained programming method itself.

In the chance-constrained approach, one specifies for each of the constraints of the original problem the probability with which it is desired that this constraint be satisfied. For example, the problem in (5.1) would be transformed into

$$(5.2) \quad \text{Maximize } f(x) \quad \text{Subject to } \Pr\{Ax \leq b\} \geq \alpha, \quad x \geq 0,$$

where Pr means probability and  $\alpha$  is a column vector of probabilities, that is,  $0 \leq \alpha_i \leq 1$  for all  $i$ . The shortcoming of chance-constrained programming in the present author's view -- as well as in the views of others -- is its failure to specify what happens if the selected values of the decision variables, the elements of the vector  $x$ , do prove to be infeasible.

One can set the  $\alpha_i$ -value for a particular constraint so that it reflects the cost of violating this constraint. This is, in fact, what the chance-constrained programming writers have in mind. If violating one budget constraint would "cost" more than violating another, the  $\alpha_i$ -value for the first should be greater than that for the second. The chance that the decision taken will require more than the first budget quantity should be smaller than the probability that the second budget will be exceeded. But what does the decisionmaking unit do that  $(1-\alpha_i) \cdot 100$  per cent of the time its course of action leads to a violation of the  $i$ th restriction? The chance-constrained approach has nothing to say to this point.

Contingency plans for the decisionmaker in case of infeasibility are not explicitly taken into account. But the need for contingency plans in case constraints are violated is surely one of the central features of a situation in which a decision must be made before all the information is at hand. A method that does not concern itself with contingency plans therefore has serious deficiencies as a tool for decisionmaking in such situations.

Unfortunately, when one tries to build a programming model that incorporates contingency plans for the case in which constraint-matrix elements are stochastic, one meets with little success. The earlier quotation from Hadley's survey summarizes this unhappy situation all too well. And this is the situation that would result if one did not adopt a period-analytic approach to the problem of capital budgeting under risk. The reason is not difficult to see.

In each period within its horizon, the firm has the cash throw-off from its originally controlled resources to use for investment purposes. If some of the projects the corporation planned to undertake were to generate returns to the firm within the individual periods, these additional funds could also be used to finance investment projects. But then the  $t$ th budget constraint would be meaningful only if the coefficient of each project variable were the net input of cash required by the project in that period.<sup>13</sup> The net expenditure on a project in a period is, however, obviously the negative of the net cash inflow in the period that is attributable to the project. Since the model of capital budgeting under risk to be discussed considers the gross returns on proposed projects stochastic but gross cash outlays on projects deterministic, the net returns on projects are stochastic. But this means the net inputs of cash are stochastic and, consequently, the constraint matrix would contain random variables if projects generated returns within (as opposed to at the end of) the individual periods.

The period-analytic framework set forth earlier avoids this difficulty. With gross returns ( $a_{ti}$ 's) from projects becoming available only at the ends of periods, they are not available for use within the periods in which they are generated. Similarly, the certain returns from the firm's operations apart from the new investment program, that is, the cash throw-offs due to initially owned resources ( $C_{t-1}$ 's), are not available for use until the periods following the ones in which they are earned. The deterministic gross cash outlays ( $c_{ti}$ 's), on the

---

<sup>13</sup>Baumol and Quandt (1965, p. 321) make the same point.



other hand, occur at the starts of the periods. Hence, a project's demand on the scarce budget dollars of the firm in a particular period (its  $c_{ti}$ -value) would seem to be known with certainty in the present model. There is only one further difficulty to be considered. Once it is overcome the elements of the budget-constraints part of the constraint matrix of the programming problem under risk will be ensured to be deterministic.

This final difficulty involves the net returns from the projects, even though it is assumed that they become available at the ends of the several periods. There are two polar cases: (1) the net returns are always all made available for withdrawal from the firm or (2) they are always all reinvested in the firm. Clearly, an entire spectrum of possibilities lies between these two extremes. The cases involve various apportionments between availability for withdrawal and for reinvestment. It should be noted that the case in which net returns are not used within the investment horizon but are kept to be used by the firm in post-horizon periods can, for present purposes, be considered a subcase of (1). The crucial feature of such an arrangement, from the perspective of the present discussion, is that the net returns are not reinvested prior to the horizon period.

The model constructed here will assume that all net returns are immediately made available for withdrawal from the firm. Without this assumption the problem of stochastic elements in the constraint matrix would arise once again. If net returns were to be reinvested in the period(s) following their realization, they would serve as a supplement to the cash throw-off from initially owned resources in that (those) period(s). But this means that the coefficient of project  $i$  in the  $t$ th budget constraint should be the gross cash outlay (if any) that  $i$  requires in period  $t$  minus the net return it earned in period  $t-1$ .<sup>14</sup> It would also

<sup>14</sup>Assume, for simplicity, that the net returns earned in period  $t$  are immediately devoted to supplementing the cash throw-off in period  $t+1$ . If they are not "needed" in that period a cash carry-forward activity, as the lending activity to be included in the model here, could postpone their use until a period when the firm needed them more.

be necessary to change the arguments in the utility function in (4.4) to the  $R_t^*$  values alone, where -- it will be recalled --  $R_t^*$  is the residual in the  $t$ th period from the net earnings of the firm's initial resources after all investment outlays have been made. This would have to be done to avoid double-counting the benefits the projects' earnings provide the owners. All of the benefits would be a function of the net amounts available for withdrawal after (i) the cash throw-offs due to initial resources, (ii) the current cash outlays on new projects, and (iii) last period's net returns on new projects had been considered. That is, they would be a function of the  $R_t^*$ -variables,  $t=1, \dots, T$ , alone.

We would be in the same situation as if gross returns were allowed to accrue within the period. The  $t$ th-period gross cash outlay on the project is deterministic, but the net return from the project in the previous period is stochastic. This is true since the project's  $t-1$ st-period gross return is stochastic, while its  $t-1$ st gross outlay is deterministic. Hence, the coefficient of the project in the  $t$ th-period budget constraint would again have to be a random variable if reinvestment of net returns were permitted. The constraint matrix would contain stochastic elements and we would return to the unhappy situation described by Hadley.

By assuming, however, that net returns from investments are always made available for withdrawal and not reinvested in the firm, it follows that a project's coefficient in the  $t$ th-period budget constraint is the gross cash outlay on it in that period. Since this outlay figure is deterministic, the constraint matrix of the programming model of capital budgeting under risk will also be deterministic. The difficult problem of a stochastic constraint matrix is thus avoided.

This assumed complete removal of projects' net returns from the firm is required if the problem of capital budgeting under risk is to be tractable. But it seems, at least at first glance, that this supposition seriously inhibits the

flexibility of the model. This is not completely true. Owners' preferences between reinvestment of earnings from projects (hence, retention of earnings) and withdrawals make themselves known through the owners' utility function. If the ownership would prefer to see earnings of the investment program put to work in the firm rather than disbursed to them, this will be reflected in a higher preference for later rather than earlier withdrawals. In contrast, a higher priority for the distribution of earnings would take the form of a higher preference for near-term withdrawals than for later ones. The firm with owners who prefer reinvestment of earnings will find itself choosing projects with a more distant payoff, a form of reinvesting net earnings. The firm with an ownership preferring early distribution of earnings will emphasize proposals with earlier payoffs in its capital budget.

The assumption that net returns from new investments are not reinvested in the firm does limit the comprehensiveness of the model. But what it does, in addition, that cushions this loss in generality is to cause the desire for and act of reinvestment to appear in another form. Specifically, as the desire for reinvestment strengthens, the "average" length of time between the start of the investment program and the periods in which the selected projects start yielding positive net returns will increase.<sup>15</sup>

Within the period-analytic framework of the present model, then, the firm makes its capital-budgeting decision. It has physical projects from which to choose as well as the option in each period of borrowing (up to a set limit) or lending at fixed but divergent rates of interest. In addition, in each period the firm can plan to make available for withdrawal from the firm some of the net earnings resulting from its initial resources, that is, part of the  $C_{t-1}$ 's. It is to a fuller description of these various options that we now turn. They are the building blocks of the programming model.

---

<sup>15</sup>The use of the term "average" is intentionally vague. No relationship exists between the use of the term here and the Knight-Austrian School debate. It is simply intended as an intuitive way of summarizing the two examples given above.

#### 5.4. The Physical Investment Projects and Physical Interdependences

The potential investment projects are assumed to be  $n$  in number. One project may be building a new plant at location A, a second may be building a plant at B, a third may be pursuing a particular area of research and development, a fourth may entail replacing a given machine with a new one, and so on. Each investment proposal can be represented by a vector, similar to the description of any activity in an activity-analysis problem.

The first  $T$  elements in the  $2T$ -element vector describing the  $i$ th project are the period-by-period gross returns of cash from the project, the  $a_{ti}$ 's for  $t=1, \dots, T$ . Each of these entries is finite and nonnegative. The firm can at most lose its entire investment and can never earn an infinite amount from a project. The last  $T$  elements of the vector describing project  $i$  are the period-by-period gross cash outlays,  $c_{ti}$ ,  $t=1, \dots, T$ , on the project. These outlays are also all nonnegative. Either a project requires funds in the  $t$ th period or it does not. And, in light of the period framework taken here, it cannot generate funds at the start of the  $t$ th period for use by the firm in that period. Suppose, for example, a model with a five-period horizon is being considered. The third element in the ten-element vector describing a proposal, say, to buy a particular machine, would then indicate the gross inflow of cash (if any) in the third period resulting from that purchase. The eighth element in the project's vector would be the outflow of cash in the third period necessitated by the acquisition of the machine.

The fact that a project begun in one of the  $T$  periods of the firm's planning horizon will yield returns in post-horizon periods creates something of a problem. Some modelbuilders in capital budgeting seem to side step the issue, tacitly assuming the firm's planning horizon is long enough to encompass all returns from all projects started during it. Others discount post-horizon flows at some discount rate, usually the company's "cost of capital" or some

market rate of interest.<sup>16</sup> Both approaches seem less than fully acceptable. The latter is not completely satisfying due to the difficulties surrounding the concept of the "cost of capital" and the meaning of a market rate of interest in an imperfect-capital-market setting; the former is not fully acceptable because it is hard to envision a firm with a planning horizon long enough so that all post-horizon flows are negligible.

I shall attempt to deal with the problem by including in the Tth-period gross-return figure for project *i* not only the gross cash inflow from project *i* in period *T* but also the market value of the project in period *T*. If, for example, the project is a building, the market value of the building in period *T* will be added to the cash inflow in period *T* to get its gross-return figure for that period. Problems of valuation will undoubtedly arise, as in attempting to assess the market value of a research and development program even when salary and equipment figures are available. But the error involved in measuring such items does not seem as great as -- or at least it seems no greater than -- the errors generated by methods used elsewhere. In essence, the market is left to complete the task the firm itself cannot really adequately perform.

It is not being suggested that this approach is the correct way to cope with post-horizon flows from the projects being considered for acceptance in the present program. In fact, this procedure may encounter the same problems that confront the approach relying on the firm's "cost of capital." This can occur because the market value of a project in period *T* is essentially the "market's" valuation of the discounted present value of post-horizon cash flows from the project. If the "market" finds settling upon a set of discount factors as difficult a job as the firm does, the present suggestion comes to little more than the selection of an arbitrary discount rate somewhere between the market's lending and borrowing rates for use in evaluating the project's value as of period *T*. It then becomes

---

<sup>16</sup> See Baumol and Quandt (1965) as an example of the former and Weingartner (1963) as an example of the latter.

equivalent to the second approach cited above. The alternative method suggested here of incrementing the  $T$ th-period gross return by the project's market value in period  $T$  is simply presented as an alternative approach to an unresolved question. It is an alternative which may allow more clear-cut evaluation of post-horizon flows in at least some cases. In any event, either the selection of an arbitrary "cost of capital" figure or the present suggestion stands on firmer ground than assuming all post-horizon flows are negligible.

The investment proposals considered in this study are discrete indivisible alternatives. A given project is either accepted in full or it is completely rejected. The firm does not have the option of buying  $3/4$  of a new machine or  $2/5$  of a research and development program. It either buys or does not buy the machine, and it either pursues or does not pursue the research and development program. At any one time it may have the option of engaging in research and development programs of different scales or building new plants of different sizes. Each of these ventures, which are identical except for size, will be considered discrete indivisible alternatives. As some support for this view, consider Turvey's statement about one of the basic facts of public investment choices. He writes, "there is always a small number of projects to be considered. Even in the rare cases where the size of a project is highly variable, no one will bother to work out costs and returns for more than two or three alternative sizes of the scheme. In consequence, it is no use talking in terms of marginal analysis."<sup>17</sup> More will have to be said about the interdependence between several projects of the same type but different size. This will be attended to presently.

In short, then, since  $y_i$  indicates the extent to which the  $i$ th potential project is accepted,  $y_i$  is an integer-valued variable. Moreover, if two projects identical in all respects are proposed simultaneously, they will be considered

<sup>17</sup>Turvey (1963, p. 94).

distinct projects. This means each  $y_i$  is not only restricted to be integer-valued, but it is further required to be a zero-one variable. The variable  $y_i$  has value zero if project  $i$  is rejected and it has value one if project  $i$  is accepted. This restriction is necessary in order to ensure that if a proposal is "best," in some objective-function sense, the optimal solution does not call for accepting it several times when multiples of the project do not really make sense. For instance, just because building a new plant at location A may be "best" among all proposals, it does not follow that two plants at A are even better.

Alternatively, all  $y_i$ -variables could have been required to take on integral values and a further restriction imposed on the subset of  $y_i$ -variables representing all-or-nothing decisions to force them to be zero or one. The approach taken here is preferable for present purposes because the capital-investment proposals for which this model is intended are basically large ventures. Hence, in most cases the projects will be of the yes-no type. An ability or a desire to accept a proposal more than once will be the exception rather than the rule.

The possibility that several proposals might be identical except for scale introduces the question of physical interdependence among projects. There are two basic types of physical interrelationship among proposals: mutual exclusion and contingency. Recall that a set of projects is mutually exclusive if acceptance of any member of the set implies automatic rejection of all other members. If, for example, the firm builds a plant of one size at a location it is not going to build a second plant of a different size at that same location. In contrast, a contingent proposal is one whose acceptance depends on the simultaneous acceptance of another project or of several other projects. The firm would not purchase equipment for a new plant without first having decided to construct the plant. But the new plant might be built even if the particular machinery were not purchased -- the plant might be used for another purpose.<sup>18</sup>

---

<sup>18</sup>For a further discussion of these physical interrelationships and their possible constraint representations see Weingartner (1963, pp. 11, 32-33; and 1966, Section III).

These physical-interdependence relations constitute another set of constraints to be added to the period-by-period budgets and the zero-one restrictions on the  $y_i$ -variables. Suppose  $S$  is a set of mutually exclusive projects. The relationship of mutual exclusion can then be expressed by the constraint

$$(5.3) \quad \sum_{i \in S} y_i \leq 1 \quad ,$$

where the summation extends over all  $y_i$  for projects in the set  $S$ . Since each  $y_i$  is constrained to be either zero or one, the fact that (5.3) insists the sum of all the  $y_i$  in the set  $S$  be at most unity means that at most one  $y_i$  for  $i \in S$  can be unity. At most one member of the set of mutually exclusive projects can be accepted.

One especially interesting case of mutually exclusive projects ought to be noted. It arises when the firm has the option of starting a given physical project in different time periods. That is, it occurs when the firm has the option of postponement for at least some of its projects. Due to differences in the value of dollars in different periods, the benefit stream derived from a project depends on the period in which it is started. The cost expediency of starting it in alternative periods will also differ because of the varying degree of stringency of budget constraints and borrowing limits. By considering the same physical project begun at different dates as different projects, the time element of investment planning, the importance of calendar time, is introduced. Marglin has made a very strong and convincing argument about the importance of calendar time in the investment decision. The way the present model takes account of the postponability option seems to fulfill the need Marglin sees. He writes, "we must substitute a dynamic decision framework, in which the construction date is a choice variable, for the static framework, in which only the yes or no question



of the desirability of present construction must be answered."<sup>19</sup> Clearly, proposals to undertake the same physical project in different periods constitute a set of mutually exclusive projects; acceptance of one of them forces rejection of all the others.

Turning to the case of contingent projects, suppose acceptance of project A is contingent upon project B's acceptance. This contingency relationship is expressed by the restriction

$$(5.4) \quad y_A - y_B \leq 0 \quad .$$

The effect of the constraint is clearly seen if it is written as  $y_A \leq y_B$ . Variable  $y_B$  can be either zero or one. If it is zero, then  $y_A$  is necessarily zero -- proposal A cannot be accepted unless B is accepted. If, instead,  $y_B$  is unity, then one has  $y_A \leq 1$ , which simply repeats the upper bound of the previous zero-one restriction on  $y_A$ . Note that the contingent-project relationship in no way constrains the independent project B.

The presence of (physical) contingency relationships can be treated in yet another way. One constructs compound projects with each compound project consisting of the contingent proposal and the proposal (or set of proposals) upon which its acceptance hinges. "Thus a compound project is characterized by the algebraic sum of the payoffs and costs of the component projects plus, perhaps, an 'interaction' term."<sup>20</sup> A particular project's contingent status can then be represented by making the compound project of which it is a part mutually exclusive with each of the independent projects upon which it depends.

<sup>19</sup>Marglin (1963, p. 11). See Chapter 2 of his book for the full presentation of his argument.

<sup>20</sup>Weingartner (1966, p. 492). This is also one of the few places in the capital-budgeting literature where cash-flow interactions between projects have been noted. Notice, though, that the presence of the interaction is here intimately bound up with the physical relationship existing between the projects. It was pointed out earlier that cash-flow interdependence can exist apart from any physical counterparts.

### 5.5. Cash-Flow Interactions Among Projects

As was emphasized earlier, however, physical interrelationships are not the only direct interactions that exist among projects. Interdependence among projects can also take the form of interaction in their cash flows. Here, as has been indicated, the returns from and/or the outlays on several projects undertaken simultaneously diverge from the algebraic sum of their individual returns and outlays, respectively. In his survey of capital-budgeting methods actually used by firms Istvan notes that interrelations among different proposed ventures can make the budgeting process extremely difficult. The interactions make it infeasible to isolate the advantages of particular projects.<sup>21</sup> Since a programming approach is advocated here, the concern is not with isolating the advantages of individual proposals. Instead, the question to be answered is how these cash-flow interactions can be incorporated within the programming framework. There are several alternatives.<sup>22</sup>

The first way cash-flow interactions can be built into the model is similar to Weingartner's compound-project approach to contingency relationships. Suppose the projects whose cash flows interact are proposals A and B. The variable  $y_A$  indicates whether or not project A is accepted and  $y_B$  shows the status of project B in the investment program. Create a new compound project with cash inflows equal to the net combined returns when projects A and B are undertaken simultaneously and cash outlays equal to the net combined expenditures when the projects are pursued jointly. Note that "net" here refers to net of interactions between the proposals and not to a distinction between net and gross figures for individual projects. The returns of individual projects being measured are still gross cash inflows, the costs are still gross cash outlays.

---

<sup>21</sup>Istvan (1961, p. 51).

<sup>22</sup>In the discussion of cash-flow interactions, consideration is confined to interactions between two projects. The alternative methods for coping with interdependent cash outlays and returns can be extended easily to higher-order interactions.

Let  $x_{AB}^*$  indicate the acceptance or rejection of this new compound project AB. Accepting the compound project,  $x_{AB}^* = 1$ , means accepting both projects A and B and taking account of their cash-flow interactions with one another. Rejecting the compound project,  $x_{AB}^* = 0$ , means that not both A and B are included in the investment program. Either one of them may be accepted as an individual project. But the two proposals can be accepted concurrently only by accepting the compound project whose activity vector describes the net combined cash flows, returns and outlays, from such joint acceptance.

Three additional constraints are required to complete the definition of this compound-project variable. They are

$$(5.5) \quad y_A + y_B + x_{AB}^* \leq 1, \quad x_{AB}^* \geq 0, \quad \text{and } x_{AB}^* \text{ integer.}$$

Since  $y_A$  and  $y_B$  are zero-one variables, the constraints in (5.5) indicate that only one of the three proposals A, B, and AB can be accepted in any particular capital program. The firm can, then, accept either project A or project B individually. But, if it wants to pursue A and B concurrently it cannot accept them on an individual basis for (5.5) implies one cannot have  $y_A + y_B = 2$ . Instead, it must accept the compound project AB in order to pursue A and B in the same budget. With the compound project accepted, it is also true that neither A nor B can be accepted in addition as separate proposals. This follows because with  $x_{AB}^* = 1$ , the first constraint in (5.5) implies  $y_A + y_B = 0$  and hence  $y_A = y_B = 0$ . But this is appropriate since both A and B are accepted by the acceptance of AB.

This use of compound projects to capture cash interactions differs from Weingartner's use of them in considering contingency relationships. The difference rests with the fact that the contingency relationship imposes no restraints on the independent projects while the compound-project approach to cash interactions certainly does. Suppose that  $x_{D:AB}^*$  represents a compound project

consisting of contingent project D and the two independent projects A and B upon which its acceptance hinges. The constraint set representing this relationship would be

$$(5.6) \quad y_A + x_{D:AB}^* \leq 1, \quad y_B + x_{D:AB}^* \leq 1, \quad x_{D:AB}^* \geq 0, \quad x_{D:AB}^* \text{ integer.}$$

As in the case of cash interactions, the compound-project and the individual-project variables cannot all be unity simultaneously. The compound project cannot be undertaken simultaneously with either or both of the independent projects. If either or both of  $y_A$  and  $y_B$  are unity,  $x_{D:AB}^*$  must equal zero. Similarly, if  $x_{D:AB}^*$  is unity -- the compound project including A and B is accepted -- then neither A nor B can be accepted individually just as in the cash-flow interactions case. The difference rests with the fact that in the case of the contingency relationship both independent projects can be accepted if the compound project is not. With  $x_{D:AB}^* = 0$ , the constraints in (5.6) imply only that  $y_A \leq 1$  and  $y_B \leq 1$ , constraints already included in the model. But, in the case of cash-flow interactions, the two projects cannot be accepted simultaneously except as part of the compound project. With  $x_{AB}^* = 0$ , the constraints in (5.5) preclude that both  $y_A$  and  $y_B$  equal unity. The constraints in (5.6) do not exclude this possibility.

Two other ways for incorporating cash-flow interactions among projects into the model also entail creating a set of artificial projects. These are, however, of a different type than the ones created in the compound-project approach. Suppose, again, that the projects with interacting cash flows are A and B. Create a new "increment project" with a vector of cash inflows equal to the vector of differences between the net combined returns when projects A and B are undertaken simultaneously and the sum of the individual returns of the projects. The stream of cash outlays of this increment project is the set of

differences between the net combined expenditures when the two projects are pursued jointly and the sums of their individual cash outlays. The relationship between this increment project and the compound project defined earlier is straightforward. If  $c_{t,AB}^*$  and  $a_{t,AB}^*$  denote the gross cash outlay and gross return, respectively, on the compound project in period  $t$ , and  $c_{t,AB}$  and  $a_{t,AB}$  the outflow and inflow, respectively, due to the increment project in period  $t$ , then,

$$(5.7) \quad \left\{ \begin{array}{l} c_{t,AB}^* = c_{tA} + c_{tB} + c_{t,AB} \quad \text{and} \\ a_{t,AB}^* = a_{tA} + a_{tB} + a_{t,AB} \end{array} \right.$$

For the interacting pair of projects A and B, define the variable  $x_{AB}$  that will indicate whether the increment project is accepted or rejected. The increment project is to be accepted if and only if both proposals A and B are included in the capital budget. Hence,  $x_{AB}$  is one if and only if both  $y_A$  and  $y_B$  are unity. If both projects are accepted and only if both projects are accepted does the firm gain or lose from their cash-flow interactions.

The two methods using increment projects for coping with cash-flow interactions differ in the constraints used to define the  $x_{AB}$ -variable. The first method imposes the following constraints:<sup>23</sup>

$$(5.8) \quad \left\{ \begin{array}{l} -y_A - y_B + 2x_{AB} \leq 0 \\ y_A + y_B - x_{AB} \leq 1 \\ x_{AB} \geq 0, \quad x_{AB} \text{ integer} \end{array} \right.$$

If neither project is accepted,  $y_A = y_B = 0$ , the constraints become  $x_{AB} \leq 0$ ,  $x_{AB} \geq -1$ ,  $x_{AB} \geq 0$  which imply  $x_{AB} = 0$ . If one project is accepted but the

<sup>23</sup>This method was developed independently by W. I. Zangwill. For another application of this approach see Zangwill (1965, p. 33).

other rejected,  $y_A = 1$  and  $y_B = 0$  or  $y_A = 0$  and  $y_B = 1$ , the restrictions reduce to  $x_{AB} \leq 1/2$ ,  $x_{AB} \geq 0$ , and  $x_{AB} \geq 0$ . Together with the requirement that  $x_{AB}$  be an integer, these constraints force  $x_{AB}$  to zero again. Lastly, suppose both projects are included in the capital program. Then  $y_A = y_B = 1$  and the constraint set becomes  $x_{AB} \leq 1$ ,  $x_{AB} \geq 1$ ,  $x_{AB} \geq 0$ . The increment project must be accepted:  $x_{AB} = 1$ .

The second of the two methods using the increment-project approach replaces the first structural constraint and the integer requirement on  $x_{AB}$  in (5.8) with two different structural constraints. In this approach, the increment project's value is dictated by the following constraint set:

$$(5.9) \quad \left\{ \begin{array}{l} -y_A + x_{AB} \leq 0 \\ -y_B + x_{AB} \leq 0 \\ y_A + y_B - x_{AB} \leq 1 \\ x_{AB} \geq 0 \end{array} \right.$$

The reader can verify that the implications of this constraint set for the cases in which both projects are accepted or both rejected are identical with those of the restrictions in (5.8). Note, however, that the integer restriction on  $x_{AB}$  is no longer needed. With constraint set (5.8) the integer requirement was needed only when one project was accepted and the other rejected. Examining this case using (5.9) instead, suppose project A is accepted but project B is rejected. The constraints in (5.9) become  $x_{AB} \leq 1$ ,  $x_{AB} \leq 0$ , and  $x_{AB} \geq 0$ . In short, due to the last two constraints  $x_{AB}$  must be zero.

Each of the increment projects created in these last two approaches to cash-flow interactions is similar to a contingent project. An increment project's acceptance depends on the acceptance of the projects whose interactions it is capturing. But the artificial projects represented by the  $x_{ij}$ -variables differ from ordinary contingent projects. While a contingent project may be accepted

or rejected if the project(s) upon which it depends is (are) accepted, an increment project represented by an  $x_{ij}$ -variable must be accepted if the proposals upon which it is contingent are in the budget. In terms of constraints, if AB were a contingent project rather than an increment project, it would be subject to all the restrictions in (5.8) or (5.9) except the second one in (5.8) or the third one in (5.9), namely,  $x_{AB} \geq y_A + y_B - 1$ .

The three methods for incorporating cash-flow interactions presented thus far have one other important feature in common, besides correctly recording these interrelationships. Each of them permits the return and cost of the investment program in any period  $t$  to be written as the sum of the  $t$ th-period returns and costs of the individual projects, considering both real and artificial projects. We continue, for reasons of expositional simplicity, to consider only second-order interactions between projects' cash flows. Then, denoting the  $i$ th real-project variable by  $y_i$  and the  $ij$ th artificial-project variable that captures the cash-flow interactions between real projects  $i$  and  $j$  by  $x_{ij}^*$  or  $x_{ij}$  (depending on which approach one uses), these methods enable one to write:<sup>24</sup>

$$(5.10) \left\{ \begin{array}{l} \text{Return from the investment program in period } t \text{ equals} \\ \\ \sum_{i=1}^n a_{ti}y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n a_{tij}^* x_{ij}^* \quad \text{if (5.5) is used, or} \\ \\ \sum_{i=1}^n a_{ti}y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n a_{tij} x_{ij} \quad \text{if (5.8) or (5.9) is used.} \end{array} \right.$$

<sup>24</sup>Alternatively, all physical-project variables could be denoted  $y_i$ , whether real or artificial. The first  $n$  of these would then be the real projects, the last  $N-n$  the artificial ones. Compare, for example, the expressions that follow in (5.10) with the expression in equation (4.3) for the  $t$ th-period return from

the investment program, namely,  $\sum_{i=1}^N a_{ti}y_i$ .

$$(5.11) \left\{ \begin{array}{l} \text{Cost of the investment program in period } t \text{ equals} \\ \\ \sum_{i=1}^n c_{ti} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n c_{tij}^* x_{ij}^* \quad \text{if (5.5) is used, or} \\ \\ \sum_{i=1}^n c_{ti} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n c_{tij} x_{ij} \quad \text{if (5.8) or (5.9) is used.} \end{array} \right.$$

It is also possible, of course, to capture the cash-flow interactions in a nonlinear way. For the case of second-order interactions between projects, one could write

$$(5.12) \left\{ \begin{array}{l} \text{Return from the investment program in period } t \text{ equals} \\ \\ \sum_{i=1}^n a_{ti} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n a_{tij} y_i y_j \quad \text{and} \\ \\ \text{Cost of the investment program in period } t \text{ equals} \\ \\ \sum_{i=1}^n c_{ti} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n c_{tij} y_i y_j \end{array} \right.$$

This is, in fact, the way one of the few authors who have stressed cash-flow interdependence and not just physical interrelationships between projects, S. Reiter, formulated the problem.<sup>25</sup>

Reiter's model, however, only considers payoff interactions between projects in the absence of budget constraints. He presents a systematic solution procedure -- the steepest-ascent-one-point-move algorithm -- developed by G. R. Sherman and

---

<sup>25</sup>Reiter (1963).



himself for such models.<sup>26</sup> As they indicate, this procedure does not necessarily locate the globally optimal investment program. The method proceeds by finding a set of "locally optimal" programs, where "locally optimal" is defined in a special way as shall be seen presently. One starts with a random list of projects and changes the list by that single inclusion of a new project or exclusion of an already listed project that increases the payoff most. This one-project-move process is continued until no further increase can be effected by such changes: a "local optimum" is at hand.

This local optimization procedure is repeated using random starting points until a predetermined stopping rule<sup>27</sup> indicates that the expected increase in total payoff from another iteration is less than the cost of that iteration. The local optimum with the highest total payoff is then chosen as the solution. To repeat, the procedure does not guarantee that the global maximum will be found, but only that a "good" local optimum will be found and that the method converges to the optimal program in a finite number of steps with probability one.

More recently, Reiter and D. B. Rice have investigated the application of this discrete optimizing procedure to more general linear and nonlinear integer programming problems.<sup>28</sup> In particular, they have programmed the procedure for the class of problems in which variables are integer-valued and in which non-homogeneous quadratic forms (including the degenerate case--linear forms) appear in both the objective function and the constraints. Their method can be modified to apply to mixed-integer programming problems and to problems in which the objective function is not concave, although they do not report results on such problems. This extension of the discrete optimizing method is important for the present discussion

---

<sup>26</sup>Reiter (1963); Reiter and Sherman (1962 and 1965). Also see Weingartner (1966, pp. 494-497).

<sup>27</sup>Reiter and Sherman (1965) present alternative stopping rules.

<sup>28</sup>Reiter and Rice (1966).

because when both return and outlay interactions are taken into account with nonlinear forms, both objective function and constraint set will, as the expressions in (5.12) show, involve nonlinearities. Moreover, it is important to the present discussion that their method can be extended to the mixed-variables case because, as shall be seen shortly, the capital-budgeting model set out here is, in fact, a mixed-integer programming problem.

The experiments Reiter and Rice have performed have been encouraging. They report their experience with all-integer problems in which (i) both objective function and constraints were all linear, (ii) both objective function and constraints contained quadratic forms, and (iii) the objective function was quadratic but all constraints were linear. "The results we have gotten with our 39 test problems suggest very strongly that this solution procedure is a practical, useful way of closely approximating the solutions to programming problems of the type indicated."<sup>29</sup> Again, this method gives no pretense to locating a global optimum. What it finds is a "good" or, if you will, "very good" local optimum.

This good approximation may, in fact, be the best that one can hope for at the present time when the problem involves a quadratic objective function, quadratic constraints, and at least some integer variables. To the best of my knowledge there does not exist a theoretical procedure -- nor a computer code -- for locating the global optimum of an integer or mixed-integer programming problem where the integer parts of the objective function and constraint set are quadratic forms. For this reason I will not use the nonlinear method of accounting for cash-flow interactions. With no theoretical procedure available for solving the problem when such a formulation is used, little (if anything) useful could be said about the model beyond simply stating it.

There still remains a choice to be made among the three methods presented earlier that capture cash-flow interactions and allow the return and cost of the

---

<sup>29</sup>Ibid., p. 829. Italics are mine.

capital budget in any particular period to be expressed as a linear form.

There exist two major grounds upon which to make the decision. The first is one of computational ease, the second one of theoretical usefulness.

In terms of computation, it is important to note that no matter which of the three approaches is chosen, the capital-budgeting model will be a mixed-integer programming problem. This will become clear when the borrowing, lending, and withdrawal activities are discussed shortly. Each of the methods for taking account of cash-flow interactions being considered creates the same number of variables, namely, one for each interacting pair of projects. All three also impose a nonnegativity constraint on the newly created variable.

The methods given by (5.5) and (5.8), however, create an integer variable while the approach expressed in constraint set (5.9) does not impose this discreteness property on the  $x_{ij}$ -variable. Of the two requiring that the new variable be integral, the method in (5.5) involves only one structural integer constraint while the one in (5.8) entails two added structural restrictions in integer variables. Since computational difficulties and costs increase with the size of the constraint set and since the constraint in (5.5) is no more "complicated" than either of the constraints in (5.8), if one were to create a new integer-valued variable, the compound-project approach to cash-flow interactions would definitely be used.

There remains, then, the computational question of whether to create a new integer variable with its one additional structural constraint or to use the increment-project method, define a new continuous variable, and add the required three additional restrictions. The answer cannot be given a priori but depends, instead, on the particular algorithm used to solve the problem and may, even then, depend on the structure of the specific problem involved. The solution procedure that will be used in this study is a decomposition or

partitioning procedure due to J. F. Benders.<sup>30</sup> It will be described in detail in Chapter 8. For the present it suffices to say that in the procedure one alternates between (i) solving integer programs each with the same number of variables but with the number of constraints increasing at each alternation, and (ii) solving linear programs each with the same number of variables and same number of constraints. Linear programs are generally easier to solve than integer programs. It is, therefore, most important to keep the number of integer programming problems that must be solved as small as possible and to contain the size of the integer programs that must be solved as much as possible.

The increment-project approach reduces the size of each integer program to be solved since it replaces one integer variable in the compound-project method with one continuous variable. The only possible benefit that might result from introducing an integer variable instead -- that is, from using the compound-project approach -- would be a reduction in the number of integer programs that has to be solved. It is difficult to believe that this reduction (if there is any) could be great enough to outweigh the benefit from the reduction in size of each integer problem which results from use of the increment-project approach.

The fact that the increment-project approach involves two more structural constraints than the compound-project approach is not really that detrimental to the increment-project method. This is a result of using Benders' algorithm. Due to the way Benders' method proceeds -- as will be seen in Chapter 8 -- using the compound-project approach involves adding one integer variable and one constraint to each integer program that must be solved. On the other hand, the increment-project approach entails introducing three continuous variables and one constraint for the linear subproblem. Briefly, the increment-project approach has this effect because the linear subproblem

<sup>30</sup>Benders (1962).

part of Benders' method involves solving a linear program that is dual to one containing the constraints of the type in (5.9), with the values of the integer variables appearing in those constraints given. That is, it entails solving a linear program the dual of which would contain, for example, the constraints in (5.9) with  $y_A$  and  $y_B$  each equal to either zero or one -- the only variable in the constraints in (5.9) then being  $x_{AB}$ .

Using the increment-project approach thus increases the size of each linear program while using the compound-project method would increase the size of each integer program. It is difficult to say, a priori, which increase is more costly computationally since the former involves three more variables and one more constraint for a linear problem while the latter involves one more variable and one more constraint for an integer problem. The test of computational experience must tell the tale. If forced to express an a priori "hunch", I would expect the increment-project approach in (5.9) to perform better given the state of the integer and linear programming arts.

Theoretical or analytical considerations also support the case of the increment-project approach. As was discussed earlier, capital budgeting has been studied in a programming framework before. One of the interesting aspects of the capital-budgeting decision that has been relatively neglected before and that this study can hope to bring to light is the role of cash-flow interactions in this decision. These include both the deterministic cash-flow interactions discussed in the present chapter and the stochastic ones to be discussed in later chapters. For this purpose it is best to have the net value of the cash interactions exhibited as clearly as possible. But this certainly favors the use of the increment-project approach.

The choice then becomes one between the path indicated in (5.8) where an integer variable is created or the one in (5.9) where the  $x_{ij}$ -variable is continuous. The two of these can serve equally well on grounds of theoretical

usefulness. If, however, the approach in (5.9) is actually computationally superior to the compound-project approach, it certainly is computationally far superior to the integer-variable increment-project alternative since the latter involves not one but two new structural constraints in integers. And, even if the method in (5.9) is not computationally superior to the compound-project approach, it is still highly likely that it will perform better than the increment-project approach in (5.8). The latter involves adding two constraints to each integer subproblem as well as one integer variable for each  $i, j$  pair. It seems highly probable that the resulting performance will compare unfavorably with the method in (5.9) that adds three continuous variables and one constraint, for each  $i, j$  pair, to each linear subproblem.

In conclusion, it appears probable that the approach to cash-flow interactions that uses a continuous-valued increment project will outperform both of the other methods computationally. It seems highly improbable that the competing increment-project approach could provide it any challenge. Its strength relative to the compound-project approach is less clear-cut. For the purpose of theoretical economic analysis, the approach using a continuous-valued increment project is clearly superior to the compound-project approach and is as good as the other incremental-project method. As a result of these two considerations, the model presented here and used in what follows will employ the definition of the increment project in (5.9) to capture cash-flow interactions among projects.

#### 5.6. The Capital-Market Opportunities

Having completed the discussion of the firm's productive investment opportunities, consider now the opportunities the imperfect capital market offers it. The firm can enter into one-period contracts as either a lender or a borrower. The loans it makes are divided into two classes -- those used to provide returns for immediate withdrawal by the owners and those used to move

corporate funds through time. We will return to the former type of loan shortly when the cash-withdrawal or cash-removal activities are discussed. The second type of loan is simply a cash carry-forward activity that provides a return. In fact, the existence of this type of loan obviates the need for a purely internal profitless budget deferral or cash carry-over activity since no one would carry funds forward without return when a profitable alternative as such lending exists.

If the firm lends  $l_t$  dollars in period  $t$ , where  $l_t$  denotes the amount of  $t$ th-period lending the firm does to smooth its cash availabilities over time, it makes this amount available to the borrower at the start of the period. At the end of period  $t$ , it receives from the borrower the amount  $(1+r_L)l_t$  where  $r_L$  is the rate of interest on money the firm lends. This sum is then available for use in period  $t+1$ . Note that since such lending does not have immediate disbursement as its goal, neither the principal nor the interest from such loans appears in the objective function. The quantities  $(1+r_L)l_t$  for  $t=1, \dots, T-1$  simply act as supplements to the budget of the immediately following period. In addition, in the present model with its  $T$ -period horizon, the firm would never make such loans in the horizon period  $T$  since the owners would never benefit from such a loan.

As for borrowing contracts, if the firm borrows  $b_t$  dollars in period  $t$  it receives this amount at the start of the period. On the last day of the period, when it is also receiving its  $t$ th period cash throw-off from its original assets, the firm must pay its creditors  $(1+r_B)b_t$  dollars. The symbol  $r_B$  denotes the rate of interest the corporation must pay on its borrowing and, by assumption,  $r_B$  exceeds  $r_L$ .<sup>31</sup> Since it must repay the principal and interest on the last day of period  $t$ , what happens is the firm reduces the amount it has internally

---

<sup>31</sup>Recall that if  $r_L > r_B$ , the firm could make an infinite amount of money by borrowing and lending simultaneously.

available for investment in period  $t+1$  by the amount  $(1+r_B)b_t$ . It repays the principal and interest on a  $t$ th-period loan out of the cash throw-off  $C_t$ . Consequently, the  $(1+r_B)b_t$  quantities, for  $t=1, \dots, T-1$ , never appear explicitly in the utility function of the model either.

The reason for stopping at the  $T-1$ st period in the case of borrowing differs from the reason there is no  $T$ th-period  $l_t$ -variable. In this instance it is because the model, stopping as it does at the horizon, allows no possibility for repayment of a loan made in the horizon period. To avoid the possibility that the firm borrows unlimited quantities in that period that it would not have to consider repaying, the option of borrowing in the horizon period is just completely removed.

The amount the firm can lend in a period is only limited -- at the uppermost point -- by the cash throw-off it received in the previous period from its initially controlled assets. Its borrowing is, however, limited by the imperfect capital market in which it operates. Whether the limit derives from a self-imposed leverage constraint, a leverage constraint induced by the tests bankers and investment analysts use in evaluating firms, trade-credit limitations, or one of the other reasons given in Section 3.3, the firm faces a borrowing ceiling in each period. This set of constraints is easily represented. Denoting the upper limit on the firm's borrowing in period  $t$  by  $B_t$ , they are

$$(5.13) \quad b_t \leq B_t, \quad t=1, \dots, T-1.$$

It is important to note that borrowing and lending are both continuous-valued activities. The variable  $l_t$  can take on any value from zero to  $C_{t-1}$  while  $b_t$  can assume any nonnegative value less than  $B_t$ . Loans are assumed to be made in finely divisible units, dollars, not in discrete indivisible amounts or packages of money. Hence, no matter what method of accounting for cash-flow interactions one used, the capital-budgeting model with the borrowing and



lending options described here would be a mixed-integer programming problem, containing discrete and continuous variables.

### 5.7. The Cash-Withdrawal Activity and the Objective Function

In addition to using the cash throw-off it has available in a given period for outlays on projects and for loans, the firm can also distribute part of these internally generated funds to its owners. It can also borrow in a given period in order to increase the amount available for disbursement within that period. Since the loan will have to be paid back in the future -- not necessarily in the next period since the borrowing contracts are de facto renewable -- any borrowing for present withdrawal will reduce the amount potentially available for future distribution. Whether or not such borrowing is desirable in a given period will, of course, depend on the owners' utility function, the productive investments available to the firm, the return it can get by lending, and the anticipated cash throw-offs from its initially controlled resources.

At the beginning of each period, then, the corporation will set aside a nonnegative sum of money,  $R_t$  dollars, for removal or withdrawal from the firm. This is the amount it removes from its cash throw-off of the previous period,  $C_{t-1}$ , to disburse to its owners. But the disbursement is not made until the end of the tth period when the earnings from the new investment program are also distributed. Since the firm can earn a return of  $r_L$  by lending these dollars, it would be wasteful to allow the amount set aside for withdrawal to remain idle during the period. Hence, at the start of period  $t$  the firm allocates  $R_t$  dollars for withdrawal purposes. It lends this money during the tth period and has available at the end of that period  $(1+r_L)R_t$  dollars for removal from the firm in the form of payments to the owners. Clearly, what this means is that if the firm actually wants to disburse  $R_t^*$  dollars<sup>32</sup> at the end of the

---

<sup>32</sup>The use of the same notation here and in equation (4.3), where  $R_t^*$  was the amount available for withdrawal in period  $t$  before the period-analytic view was introduced, is intentional.

tth period, it sets aside  $R_t = \frac{R_t^*}{1+r_L}$  dollars at the start of the period. The withdrawal activity makes a demand of  $R_t$  dollars on the firm's tth-period funds and makes available  $(1+r_L)R_t$  dollars for distribution to the owners at the end of that period.

With the withdrawal-of-funds activity specified, it is possible to state precisely the objective function of the capital-budgeting problem in the present period-analytic model. Recall that the objective is to maximize the utility function

$$(5.14) \quad U = U(W_1, W_2, \dots, W_T)$$

where  $W_t$  is the total amount available for withdrawal from the firm in the tth period. Within the present period-analytic framework each  $W_t$  is equal to

$$(5.15) \quad W_t = \sum_{i=1}^n a_{ti} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n a_{tij} x_{ij} + (1+r_L)R_t .$$

This formulation uses, as was noted earlier, the increment-project approach to cash-flow interactions given in (5.9). The capital-budgeting problem thus has the following objective function:

$$(5.16) \quad \text{Maximize } U = U \left( \sum_{i=1}^n a_{li} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n a_{lij} x_{ij} + (1+r_L)R_1, \dots, \right. \\ \left. \sum_{i=1}^n a_{Ti} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n a_{Tij} x_{ij} + (1+r_L)R_T \right) .$$

The presence of two different types of lending activity ( $l_t$  and  $R_t$ ) requires some further explanation. The position taken here is that the borrowing and lending activities,  $b_t$  and  $l_t$ , are used simply to smooth the firm's cash availabilities through time. The firm can increase its available funds in

earlier periods at a cost of  $r_B$  per dollar for a one-period shift forward. It can move funds from the present to the future with the added return of  $r_L$  per dollar for a one-period change backward in time. The borrowing and lending possibilities the firm has in the  $b_t$ - and  $l_t$ -variables are, then, activities different from the physical investment proposals. They are not explicitly included in the objective function's arguments but rather simply move funds through time alleviating tight budget or borrowing constraints by shifting funds from periods in which the corporation's financial position is less strained.

The withdrawal activity, on the other hand, is more in the spirit of a regular physical project. Each period's (nonnegative) withdrawal makes a demand on that period's budget and provides an increment to the objective-function argument for that period. While the lending activity,  $l_t$ , is simply used to shift funds back in time, the withdrawal or removal activity,  $R_t$ , is used to disburse money to owners now and to earn a money-market return for them now.

In one major respect, however, the withdrawal activity differs from the physical projects in the model. As is the case with the  $b_t$ - and  $l_t$ -variables, each removal-activity variable  $R_t$  is a continuous-valued, nonnegative variable. Disbursements are made in dollars, as finely divided as one wants, and not in discrete amounts. Hence, as soon as a withdrawal activity of the type introduced here is included in a model with indivisible projects, the model takes the form of a mixed-integer program.

The continuous nature of the lending and withdrawal activities has one further extremely important effect. Because they are continuous-valued variables each budget constraint will actually hold as an equation, not as an inequality, in the optimal solution. With a return available by lending -- either for present distribution or future use -- none of the firm's funds will ever rest idle at the optimum. They will be absorbed either in an  $l_t$ -variable to move present funds to the future or in an  $R_t$ -variable to increase the owners' present return.

### 5.8. Completing the Constraint Set

The components of the T budget constraints have now been fully described. This budget portion of the constraint set can now be presented. It consists of the following restrictions:

$$(5.17) \left\{ \begin{array}{l} \sum_{i=1}^n c_{1i} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n c_{tij} x_{ij} + l_1 - b_1 + R_1 \leq C_0 \\ \\ \sum_{i=1}^n c_{ti} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n c_{tij} x_{ij} - (1+r_L)l_{t-1} + l_t + (1+r_B)b_{t-1} - b_t + R_t \leq C_{t-1} \\ \\ \sum_{i=1}^n c_{Ti} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n c_{Tij} x_{ij} - (1+r_L)l_{T-1} + (1+r_B)b_{T-1} + R_T \leq C_{T-1} \end{array} \right. \quad \text{for } t=2, \dots, T-1$$

The programming model of capital budgeting is almost complete. What remains to be included are any restrictions on the firm's project selection that exist in addition to the financial constraints and the physical-interdependence constraints. As the discussion in Section 3.4 indicated, these may arise in the form of manpower or personnel budgets, material limitations, and the like. Such constraints can be incorporated easily, at least in a formal sense. Let  $D_k$  represent the total amount of the  $k$ th such limited resource available to the firm, for example, personnel trained for a specific task. If  $d_{ki}$  is the amount of this limited factor (the number of trained workers) required to undertake project  $i$ , the additional constraint is

$$(5.18) \quad \sum_{i=1}^n d_{ki} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n d_{kij} x_{ij} \leq D_k$$

One such constraint exists for each of the  $K$  limited resources.

The interaction terms are included not simply for formal completeness but because it seems likely that such interrelationships may well exist. Suppose, for example, the resource in short supply is trained construction supervisory personnel and that two of the projects contemplated are building a plant at location A and building a plant at location B. Undertaking the construction at A alone might require two supervisors while undertaking the one at B alone might require three supervisors. If, however, the two construction sites are close enough to each other, supervisory personnel might be able to divide their time between the sites and the firm could economize on the limited resource. That is,  $d_{KAB}$  would be negative, and hence the number of supervisors needed if both plants were built would be less than five. No argument is being made that such interactions are nearly as widespread in the case of resource demands as in that of cash flows, particularly cash inflows. Rather, it is just being stated that such resource interdependence is, in fact, a real possibility.

#### 5.9. The Complete Programming Model: A Summary

The programming model this chapter set out to construct is now complete. A brief view of the state of the capital-budgeting art -- the way it is done by existing enterprises -- as opposed to capital-budgeting science, showed how normative, or even utopian, the programming approach being advocated here actually is. It was, in fact, noted that some approaches that could be at best viewed as second-best or third-best to the programming method remained in the realm of normative suggestions relative to what firms actually do.

Accepting this severe limitation on the operational usefulness of the model, the chapter proceeded in the hope of making a contribution by posing the questions in a different, one would hope, better way than they had been framed before. The corporation was viewed as making investment decisions within a period framework and having productive as well as financial opportunities for

investment. In addition, it had the option of disbursing funds to its owners at the end of each period by setting aside earnings from the resources it owned before embarking on its investment program. The opportunities the firm possessed as well as the constraints within which it was forced to operate were delineated. Special attention was paid to the question of cash-flow interactions among projects, including both interactions in cash outlays and in cash returns.

The programming model of the capital-budgeting problem that emerges from the discussion is presented in equation set (5.19).

In the next two chapters, the discussion turns to the subject of budgeting capital in an environment of risk. The ultimate question to be answered is: How does the existence of risk alter the model specified in the present chapter?

$$\text{Maximize } U = U\left(\sum_{i=1}^n a_{li} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n a_{lij} x_{ij} + (1+r_L)R_1, \dots, \sum_{i=1}^n a_{Ti} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n a_{Tij} x_{ij} + (1+r_L)R_T\right)$$

Subject to:

$$\sum_{i=1}^n c_{li} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n c_{lij} x_{ij} + l_1^{-b_1 + R_1} \leq C_0$$

$$\sum_{i=1}^n c_{ti} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n c_{tij} x_{ij} - (1+r_L)l_{t-1} + l_t + (1+r_B)b_{t-1} - b_t + R_t \leq C_{t-1},$$

for  $t=2, \dots, T-1$ .

$$\sum_{i=1}^n c_{Ti} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n c_{Tij} x_{ij} - (1+r_L)l_{T-1} + (1+r_B)b_{T-1} + R_T \leq C_{T-1}$$

(5.19)

$$\sum_{i=1}^n d_{ki} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n d_{kij} x_{ij} \leq D_k \quad \text{for } k=1, \dots, K$$

$$b_t \leq B_t \quad \text{for } t=1, \dots, T-1$$

$$\sum_{i \in S} y_i \leq 1 \quad \text{for each set of mutually exclusive projects.}$$

$$y_i - y_j \leq 0 \quad \text{for each contingent project (i)-independent project (j) pair.}$$

$$\left. \begin{aligned} -y_i + x_{ij} &\leq 0 \\ -y_j + x_{ij} &\leq 0 \\ y_i + y_j - x_{ij} &\leq 1 \end{aligned} \right\} \text{for each } i, j \text{ pair with } i < j$$

$$y_i = 0 \text{ or } 1 \quad \text{for each } i.$$

$$x_{ij} \geq 0 \quad \text{for each } i, j \text{ pair with } i < j$$

$$R_t \geq 0 \quad \text{for each } t=1, \dots, T$$

$$b_t \geq 0 \quad \text{for each } t=1, \dots, T-1; \quad l_t \geq 0 \quad \text{for each } t=1, \dots, T-1$$





## CHAPTER 6

### THE CAPITAL-BUDGETING PROBLEM UNDER RISK: THE RISK ENVIRONMENT AND PREVIOUS SUGGESTIONS FOR COPING WITH IT

#### 6.1. Introduction

Previous chapters have considered two sets of questions concerning decisionmaking in a multiperiod context. First, the characterization of attitudes toward risk when the planning horizon extends over several periods of time was discussed. Second, the problem of budgeting capital over a T-period horizon was considered for a firm operating in a world of certainty but facing imperfect capital markets. The next two chapters draw these two developments together as they consider the problem of capital budgeting in a risk environment.

The ultimate goal of the next two chapters is to bring the programming model developed up to this point in closer touch with the stochastic reality of the capital-budgeting decision. The present chapter begins by describing the nature of the particular risk environment considered here. Then, it considers some of the suggestions that have appeared in the literature for coping with the presence of risk in capital-investment decisions. The approaches considered include the finite-horizon approach, the risk-discount method, sensitivity analysis, the application of chance-constrained programming, the mean-variance approach, and the time-state-preference approach. Each of these suggestions is discussed and evaluated. After this review is completed, the next chapter goes on to construct a programming model of capital budgeting under risk. As shall be seen, the bulk of the effort in building this model will be devoted to determining an

appropriate objective function for the capital-investment decision that must be made in an environment of risk.

Before continuing, it is important to reiterate one of the disclaimers made in Chapter 3 so that the scope of the discussion will be more clearly defined. There will be no attempt to consider the capital-budgeting decision as a problem of sequential decisionmaking. Only the investment opportunities the firm sees arising over the course of its T-period planning horizon are considered, and the only stochastic elements of the problem relate to the cash flows associated with these opportunities.

Capital-budgeting decisions are, however, actually more dynamic than this since investment opportunities characteristically arise randomly through time. The investment decisions must be made as the opportunities arise, within the horizon, and not just at one point in time. G. M. Kaufman accurately describes this greater degree of risk or uncertainty inherent in capital-investment planning.

In selecting among presently available investment opportunities the firm must also take into account the time, number, size, and profitability of deals that may appear in the future -- none of which may be known with certainty. As a consequence, most firms are faced with two conflicting desires: to have funds available for highly profitable investment opportunities that have not yet appeared on the investment horizon and at the same time to avoid tax and opportunity cost penalties for underinvestment.<sup>1</sup>

Ultimately, one would want to construct a programming model that takes account of both the stochastic aspects of presently available or foreseen projects and the possibility that new, profitable, unforeseen projects may appear. For the present, we satisfice and consider a programming model in which the nondeterministic nature of the environment appears only in the form of risks associated with the cash flows of presently conceived proposals.

---

<sup>1</sup>Kaufman (1963a, p. 39).

## 6.2. The Nature of the Risk Environment and the Need for a Programming Approach

---

In the model presented here, it is the gross cash returns from physical projects -- both the cash inflows they generate individually and by interacting with one another -- that involve risk for the firm. That is, the  $a_{ti}$  and  $a_{tij}$  parameters are random variables with (objectively or subjectively) known probability distributions. All the other parameters of the capital-budgeting decision are assumed to be known with certainty. The firm's anticipated cash throw-offs from its originally possessed assets (the  $C_t$ 's), the gross cash outlays on physical projects (the  $c_{ti}$ 's and  $c_{tij}$ 's), the amounts of nonfinancial limited resources (the  $D_k$ 's), the nonfinancial requirements of physical projects (the  $d_{ki}$ 's and  $d_{kij}$ 's), the borrowing limits (the  $B_t$ 's), and the interest rates ( $r_B$  and  $r_L$ ) are all considered deterministic figures.

Note that it is only the physical projects that are assumed to entail risks, and then only with respect to the gross cash inflows they generate. All three financial activities, borrowing, lending, and withdrawal, are carried on with certainty as to what their cash requirements and cash returns will be. For example, setting aside one dollar for withdrawal in period  $t$  will always require exactly one dollar of that period's budget and will always lead to the disbursement of exactly  $1+r_L$  dollars to the owners at the end of the  $t$ th period.

Some justification ought to be given for considering this particular risk environment. Any model purporting to be fully realistic would want to consider all financial parameters of the firm's capital-budgeting problem as subject to risk. The cash outlays on projects, the dollar amounts generated by the firm's ongoing operations, and the interest rates at which the firm can borrow and lend are surely not truly deterministic. A more realistic portrayal of the firm's decision would want to consider them stochastic variables. But presently available analytical tools do have their limitations.

One of these shortcomings, as discussed in the previous chapter, is the present inability to solve satisfactorily linear -- no less nonlinear -- programming problems when elements of the constraints matrix are stochastic variables. Hence if the model presented here is to be amenable to analysis -- which it obviously must be if any meaningful results are to be derived -- the elements of the constraint matrix must be deterministic. The period-analytic framework discussed in the last chapter and the assumption that all net returns from the investment program are immediately withdrawn from the firm imply that the demand a project makes on the scarce capital resources of the firm in a particular period is the gross cash outlay on the proposal in that period. This enables us to consider the gross returns on projects stochastic while simultaneously having the projects' coefficients in the T budget constraints be deterministic. Risk exists in the model in the form of stochastic returns from projects, while the assumption of nonstochastic gross cash outlays maintains the tractability of the model.

The constraint matrix of the capital-budgeting model, however, contains more than projects' demands on company funds in the several periods. Examining the constraints as formulated in (5.19), those expressing mutual exclusion or contingency relationships among projects are clearly nonstochastic. The constraints used in the definition of the artificial-project variables are likewise deterministic. But the market rates of interest, which appear in the budget constraints, and the proposals' requirements of other limited company resources might be stochastic. If they were random variables, the capital-budgeting model being developed would have few insights to provide since the constraint matrix would contain stochastic elements and the programming problem would not be solvable analytically. To avoid this undesirable possibility, the borrowing and lending rates and the scarce-resource demands of the proposals have been taken to be deterministic.

One last question remains with regard to the nature of the risk environment assumed here. Why should it be posited that the anticipated cash throw-offs and the borrowing limits faced by the firm are nonstochastic? This supposition is also, in part, a matter of tractability to analysis. There do exist methods for solving linear programming problems with stochastic requirements vectors.<sup>2</sup> But none of the papers on the subject consider the problem of solving mixed-integer programming problems when elements of the requirements vector are random variables. While a reading of the paper by Dantzig and Madansky would lead one to envision a tractable formulation of the mixed-integer problem when the density functions of the elements of the requirements vector are discrete, the case of continuous random variables warrants much less hope.<sup>3</sup>

The present study makes no attempt to extend the theory of stochastic programming to mixed-integer problems. For reasons of analytical tractability, then, the borrowing limits and anticipated cash throw-offs would have to be assumed known with certainty.

Assuming these two sets of parameters to be deterministic is actually much less bothersome than the corresponding assumption about the outlays on projects. First, if one looks at the reasons given for the possible existence of the borrowing limits, it is clear that in contrast to the returns or outlays on projects,

---

<sup>2</sup>See, for example, Dantzig and Madansky (1961); Hadley (1964, pp. 160-167); Madansky (1962); and Williams (1965, 1966).

<sup>3</sup>Dantzig and Madansky (1961, p. 170). It is important to note that when Dantzig and Madansky turn to the computational aspects of their linear programming model they look for ways to avoid solving it in its original large-scale form. If the solution of the linear programming model entails such computational complications, one need not stretch his imagination too far to realize the difficulties that would arise were the program a mixed-integer problem. In addition, the computational simplifications Dantzig and Madansky introduce, specifically, relying on the Dantzig-Wolfe decomposition principle [see Dantzig and Wolfe (1960)], could not be applied in the mixed-integer case. Hence, although one can envision an extension of the Dantzig-Madansky model to the mixed-integer case when the density functions involved are discrete, the computational problems involved with such a model are forbidding.

it is more likely that the firm will know these borrowing limits with certainty. For example, foremost among the explanations for the borrowing limits were self-imposed leverage constraints. Whereas the firm cannot tell what returns a particular project will generate, it can most probably state (at least within narrow limits) the limitation it wants to place, for leverage reasons, on its borrowing.

When the borrowing limits and cash throw-offs are not really known with certainty, in the former case perhaps because management cannot predict the vicissitudes of banker and investment-analyst tests, there exists another justification, from the point of view of theoretical insights to be gained, for taking them to be deterministic. The reason is that assuming the cash throw-offs and the borrowing limits are nonstochastic enables us to isolate the role of the risks associated with the proposed investments themselves. This is not to say that the risks involved in the firm's ongoing operations are unimportant. They may, indeed, prove to be significant. But, generally, relative to the risks involved in undertaking the new projects, neither the risks of ongoing operations nor uncertainty about the level of borrowing limits would be of major importance. Predicting the returns from new ventures and foreseeing difficulties that may arise in undertaking them are much more difficult tasks than envisioning future problems with existing activities of the firm.

This discussion of the environment in which the firm operates has stressed the motivation of analytical tractability in decisions about which parameters are stochastic and which are deterministic in the model. All of this discussion should not be allowed to leave the reader with the idea that the resulting degree of risk is minimal. In fact, stochastic returns on investment projects comprise a large part -- if not the largest part -- of the risk a firm faces when it selects a capital-budgeting program.

Confronted with this risk environment, then, how should the firm optimally select the investment projects to pursue? The correct approach to capital budgeting under risk remains -- as it was in a world of certainty -- one grounded in the methods of mathematical programming. The case presented in previous chapters for a programming model of capital budgeting is no less cogent when risk is introduced.

In fact, to the extent that the presence of risk introduces stochastic interrelationships among projects which may have been independent of one another in the deterministic model, the need for a programming approach becomes even more critical. Consider two projects that bear no physical relationship to one another: they are not mutually exclusive nor is one contingent upon the other for acceptance. Suppose, in addition, that they have no cash-flow interactions with each other when they are considered in a deterministic framework. Two new stores constructed sufficiently far away from one another could well serve as the example. The success or failure of the two projects could, however, depend on a common factor that is truly only properly regarded as a random variable. In the case of the two stores, this factor might be the demand for a given product marketed by the corporation. The returns to the two projects -- building the stores -- will then each be correlated with this random factor and hence, most likely, with one another.

These induced stochastic interrelationships are an integral part of the capital-budgeting decision when risk prevails. For the same reasons that it was observed only a programming approach could adequately cope with deterministic interdependence -- both physical and financial -- only a programming model is really up to the task of encompassing these stochastic interrelationships. To the extent that the presence of risk multiplies the degree of interdependence among projects, it thus emphasizes the need for a programming approach to capital budgeting.

The course to be followed here is modification of the programming model developed in the previous chapter so that it can cope with the risks described. Before proceeding to that model, we review the previously suggested approaches to this problem, observing the need for a new approach. The methods that have been proposed can be divided into two classes: those that are readily usable and those that basically remain in the realm of normative theoretical suggestions. Members of both sets of proposals appear to have their theoretical shortcomings, but the difficulties with the more usable rules are greater. This is unfortunate for it turns out that the most usable methods for coping with risk are, as one would expect, most widely used.<sup>4</sup>

### 6.3. An Evaluation of the More Practical Previous Suggestions for Budgeting Capital Under Risk

The primary difficulty with the first class of methods suggested for coping with risk is their lack of a sound theoretical foundation. The exact relationship between the measures suggested and the extent to which risk is present is never made clear. One can attribute this silence to the lack of any such exact relationship between the methods suggested and the degree of risk. The rules in this first class are basically intuitively appealing suggestions which, although easy to apply, often lead to incorrect decisions.

#### 6.3.A. The Finite-Horizon Approach

The first of these methods, and the worst of them, simply shortens the horizon for more risky projects. It ignores all returns from the project after an arbitrary point in time and chooses earlier cut-off points in time as the projects considered get riskier. Baumol cites the example of Federal decision-making about waterway developments where a fifty-year horizon is chosen. Any

<sup>4</sup>For evidence on this, see Baumol (1965, p. 454) and Pullara and Walker (1965, p. 406).



net returns from the development from the fifty-first year on are simply ignored.<sup>5</sup> This approach is essentially the adaptation of the payback-period criterion to the presence of risk. Using the payback period in taking account of risk was, in fact, suggested by several capital-budgeting writers who strongly opposed its use under certainty. For example, Dean wrote "It [payback] also can be useful for appraising risky investments where the rate of capital wastage is particularly hard to predict. Since payback weights near-year earnings heavily and distant earnings not at all, it contains a sort of built-in hedge against the possibility of a short economic life."<sup>6</sup>

While the implicit assumption of the horizon-curtailling approach, that the further into the future one must peer the less clearly one sees, is valid, its resolution of the problem is unacceptable. The arguments raised earlier against use of the payback criterion under certainty apply equally well against the use of the arbitrary-horizon method under risk. Its arbitrary and complete exclusion of some flows from consideration, its discrimination between dollars received on different projects in the same year, its neglect of the timing of flows within the horizon, and its lack of concern for our increasing inability to forecast as we go further out within the horizon all provide grounds for rejecting the finite-horizon approach.

### 6.3.B. The Risk-Discount Approach

A second approach to capital budgeting under risk that seems intuitively plausible at first glance involves the use of different discount rates for projects with different degrees of risk. This procedure can be used in conjunction with either the internal-rate-of-return or discounted-present-value rules for budgeting capital under certainty.<sup>7</sup> In the former case, one would

<sup>5</sup>Baumol (1965, p. 454).

<sup>6</sup>Dean (1954, p. 26). Also see Baldwin (1959, p. 102).

<sup>7</sup>For a discussion of its use in the discounted-present-value context see Bierman and Smidt (1966, pp. 322-331) and Steindl (1941), and for a discussion of its use when the internal rate of return is the figure of merit see McLean (1958).

increase the minimum acceptable rate of return (the cut-off rate) to be compared with the project's marginal efficiency of investment. If the discounted-present-value rule were used, the discount rate would be increased for risky ventures. The more risky the project, the greater would the increment in the cut-off rate or the discount rate be.

Suppose, for example, that the unique riskless market rate of interest is  $r$ . Then projects would be evaluated on the basis of their risk-adjusted discounted present values calculated from the formula

$$(6.1) \quad \sum_{t=1}^T \left( \frac{1}{1+r+\delta^i} \right)^{t-1} V_t^i .$$

Here  $V_t^i$  is the expected net cash inflow into the firm in period  $t$  resulting from undertaking project  $i$  and  $\delta^i$  is the incremental risk discount associated with project  $i$ . The riskier the project, the higher is  $\delta^i$  and the greater the reduction in the evaluation of the project's anticipated return. If it is thought to be desirable, one can use different risk factors for different periods. That is, letting  $\delta_s^i$  denote the incremental risk discount associated with project  $i$  in the  $s$ th period, the  $i$ th project's risk-adjusted discounted present value would be

$$(6.2) \quad \sum_{t=1}^T \left\{ \prod_{s=1}^{t-1} \left( \frac{1}{1+r+\delta_s^i} \right) \right\} V_t^i .$$

The risk-discount approach has several desirable features. First, the extent to which a project's anticipated return is diminished is directly related to the risk associated with it. Second, the net returns in all periods and the risks associated with the net returns occurring in all periods are considered. This stands in contrast to the arbitrary-horizon approach in which net returns

beyond the horizon are ignored and the risks associated with those occurring before the horizon are ignored. Using the risk-discount method, the fact that the accuracy of forecasts diminishes as the period forecasted about moves further away is reflected in the decrease in the coefficient of  $V_t^i$  with increases in  $t$ . Whether one uses formula (6.1) or (6.2), so long as  $\delta_s^i$  is nonnegative for all  $s$  and  $i$ , the return of the  $t+1$ st period is reduced more than the return of the  $t$ th period.

There is, however, a problem with the way our increasing inability to see further into the future is reflected by the formula in (6.1). Namely, the risk-adjusted discounted-present-value formula given there assumes that the probability that a dollar of returns actually materializes decreases by a fixed percentage each year. The formula in (6.2), which allows for different incremental risk discounts in the several periods, avoids this difficulty at the operational expense of arbitrarily having to specify not one but a vector of risk discounts.

The theoretical shortcomings of the risk-discount approach run deeper than the assumed constant proportionality of probabilities in successive periods that appears in formula (6.1). First, the method assumes that risk affects both inflows and outflows in identically the same way. Since the method discounts incremental net returns in each period, it decreases the probability of a dollar of outlay materializing by the same factor as it decreases the probability of a dollar of inflow occurring. In any particular case, though, different risks may attach to the cash returns and to the cash outlays. It will not normally be the case that the decisionmaker's uncertainty about cash flows is of the same exact degree for inflows as for outflows when he sees risks associated with both sets of flows. In the risk environment with which this study is concerned, only cash inflows are stochastic; cash outlays are all

deterministic. Hence, the risk-discount approach is definitely not suitable for use by the firm choosing investments in the environment considered here.

Second, although adding an incremental risk discount allows for cash flows occurring with different probabilities, it does not give the decisionmaker a clear picture of the risk the firm would be assuming if it were to accept the project. There is no way to separate the risk component from the timing aspect of the cash flows when the risk-discount method is used. Instead, one obtains a muddled rate-of-return or present-value figure with neither risk nor the time value of money measured well. The risk-discount method may, then, actually suppress information about the riskiness of the investments rather than aid in the evaluation of the investments.

The risk-discount approach is thus of questionable usefulness because there is no reason to believe that the probabilities with which cash flows materialize actually change over time as the method suggests they do. In order for this approach to accurately portray the risk involved in a particular project and enable separation of the risk aspect from the time-value aspect, two conditions must be met. First, the risk involved in the project's cash flows in any period must be adequately summarized by a single figure indicating the probability with which the flows will occur, and the only change in riskiness from one period to the next must be a decrease in this probability. Second, the magnitudes of cash inflows and cash outflows in any given period must be equally uncertain. These two conditions will rarely be met simultaneously.<sup>8</sup>

Two final difficulties with the risk-discount method should be recognized. The first is that the method evaluates projects one at a time, just as its certainty counterpart does. No account is taken of the interrelated risks of different projects although these risks may be of great importance, perhaps even greater in magnitude than the risks of the individual projects considered alone.

---

<sup>8</sup> For a further discussion of the shortcomings of the risk-discount approach, see Robichek and Myers (1965, pp. 79-86).

Second, as for its operational usefulness, the method is never presented with a "how-to-choose-them" manual for calculating the  $\delta_s^i$ -figures. The decision-maker is forced to rely on intuition and subjective judgment in setting the values of these parameters. Subjective judgment has been introduced into the analysis presented here with the role given to subjective probability estimates and utility functions. What is important to note is that the supposedly heuristic and easily operational approach -- the risk-discount approach -- also involves a considerable amount of subjective judgment. Moreover, the clarity and meaning of what is being judged in the two cases is very different. In the case of personal probabilities, for example, the decisionmaker is asked to state the probability with which he thinks different returns or outlays will occur. In choosing the values of each  $\delta_s^i$ -parameter, on the other hand, he is being asked to present a figure that represents some loose amalgamation of the probability with which he expects the mean value of the flow to occur and his aversion to risk. While there is a clear theoretical underpinning for personal or subjective probabilities and for utility functions, there does not exist a firm grounding for this potpourri of information contained in the risk-discount figures.

### 6.3.C. Sensitivity Analysis

The last "practical" method for considering risk in capital budgeting which will be discussed here is sensitivity analysis. The idea behind this approach is simple. The decisionmaker focuses on those parameters of the decision that are most important and also most realistically treated as being subject to risk. He then calculates the figure of merit for the project -- be it discounted present value, internal rate of return, or payback period -- assuming different values for these strategic variables. The objective of the exercise, as its name would suggest, is to see how sensitive the results are to changes in the values of these parameters. In addition, techniques known as parametric

programming algorithms exist for investigating how optimal solutions change when objective-function and requirements-vector parameters are varied, thus enabling a sensitivity analysis of linear programming models of capital budgeting.<sup>9</sup>

Sensitivity analysis seems useful as a first step in the analysis of capital budgeting under risk. It can give the decisionmaker a picture of the extent to which the stochastic elements of the environment can affect his final position. In effect, it can provide a vivid demonstration of how thick the production-possibility locus really is as a result of the presence of risk. In terms of conventional geometric analysis, sensitivity analysis sets out the decisionmaker's possibility locus and the way that locus may be shifted because of stochastic forces, but it says nothing about his indifference surfaces and the determination of an optimum. Thus, sensitivity analysis is a useful way of displaying information about the environment but it is no substitute for a full decision procedure.

#### 6.4. An Evaluation of the More Theoretical Previous Suggestions for Budgeting Capital Under Risk

Turning to the more theoretical suggestions that have appeared in the literature, it should be kept in mind that they have, by and large, remained in the realm of normative capital-budgeting procedures. Although they are superior on theoretical grounds to the methods just discussed, they are inferior operationally. They are not so easily put into use by the businessman as the arbitrary-horizon or risk-discount or sensitivity-analysis procedures. Two of these previous theoretical suggestions, however, also have theoretical shortcomings. This provides a modicum of justification for adding, as the present study does, to the stock of normative though not presently operational approaches to capital budgeting under risk.

<sup>9</sup>Gass and Saaty (1954, 1955a, 1955b). See Weingartner (1963, pp. 129-137) for examples of the application of these techniques to capital-budgeting problems considered as linear programs.

#### 6.4.A. The Chance-Constrained Programming Approach

One previous theoretical suggestion places the capital-budgeting problem in the framework of chance-constrained programming. Models of this type are presented and discussed by F. S. Hillier and B. Naslund.<sup>10</sup> In the chance-constrained programming models, the net cash inflow from potential investment project  $i$  in period  $t$ ,  $a_{ti} - c_{ti}$  in the present notation, is a random variable. In any given period, in addition to its physical-investment options, the firm can also engage in borrowing or lending on the money market, each activity taking the form of the one-period contracts described earlier. The firm has cash throw-offs, the  $C_{t-1}$ 's of the present model, being generated by the rest of its operations over the course of its investment horizon. The corporate objective is taken to be maximization of the expected value of the firm's horizon value (in Naslund's model) or maximization of the expected discounted present value of the investment program (in Hillier's model).

Recognizing the stochastic nature of future cash flows, the budget constraints the firm faces are formulated probabilistically. The  $s$ th constraint states that the accumulated result of the firm's operations -- its physical-investment program, its borrowing and lending activities, and the remainder of its operations -- through the  $s$ th period shall not be a net loss more than  $(1 - \alpha_s) \cdot 100$  per cent of the time. Alternatively, "the constraints express, that in  $\alpha \cdot 100$  per cent [where  $\alpha$  is the T-element vector of  $\alpha_s$  probabilities] of the cases, there will be enough money borrowed and generated by the rest of the business and by the investment projects, to cover the outflow of money from lending and investment activities."<sup>11</sup>

Using the notation introduced in Chapter 5, the constraint for the  $s$ th period would be

---

<sup>10</sup>Hillier (1964, Chapter 6) and Naslund (1964, Chapter 6; and 1966).

<sup>11</sup>Naslund (1964, p. 115).

$$(6.3) \left\{ \begin{aligned} & \Pr \left\{ \sum_{t=1}^s \sum_{i=1}^n (c_{ti} - a_{ti}) y_i + \sum_{t=1}^s \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n (c_{tij} - a_{tij}) x_{ij} - \sum_{t=1}^{s-1} r_L \ell_t \right. \\ & \left. + \sum_{t=1}^{s-1} r_B b_t + \ell_s - b_s \leq \sum_{t=1}^s C_{t-1} \right\} \geq \alpha_s . \end{aligned} \right.$$

Note that with the net demands on corporate funds stochastic, it can no longer be stated that at the optimum the firm's sth-period borrowing or lending will exactly compensate for the difference between its cumulated budgets and the cumulated net cash flows.<sup>12</sup>

The decisionmaker's attitude toward risk appears in these models in the choice of the elements of the  $\alpha$ -vector, not in the objective function. The objective is simply maximization of expected horizon value or expected discounted present value. The higher the  $\alpha_s$ -values are, the more certain it is the budget constraints will be fulfilled. The more averse to risk the decisionmaker is, then, the higher the set of  $\alpha_s$ -values he will select. Advocates of chance-constrained programming realize that it may be difficult for the decisionmaker to specify the  $\alpha_s$ -probabilities since he may be unaware of the effect of his choice upon the optimal solution. Naslund suggests that the use of the dual evaluators of the constraints can help the decisionmaker in this selection process. By indicating the effect on the optimal value of the objective function of raising or lowering any  $\alpha_s$  by one per cent, the dual values can help the decisionmaker reach the most desired set of  $\alpha_s$ -values in accord with his unrevealed utility function of return and risk.<sup>13</sup>

<sup>12</sup>The constraint in (6.3) is a representation of the type of constraint Naslund uses, except for the fact that in his model  $r_B = r_L$ . The capital-market imperfection he considers is the existence of borrowing limits in each period. Hillier's model is similar to the one described here except that he not only sets a probabilistic upper bound on cumulated net losses but also on the net loss in each period individually.

<sup>13</sup>See Naslund (1964, Chapter 2).



At the present time, however, there does not exist a general solution to the chance-constrained integer programming problem that has just been described. In fact, there is not available a general solution for the problem that would result if the zero-one restrictions on projects were removed, leaving only a set of unity upper-bound restrictions on the  $y_i$ 's. The course followed by those who have constructed chance-constrained programming models is to introduce a special assumption at this juncture. The random variables,  $c_{ti} - a_{ti}$ 's and  $c_{tij} - a_{tij}$ 's, are taken to be governed by a multivariate normal distribution. (For expositional simplicity, it is usually posited that they are independently distributed.) The sum in braces on the left-hand side of the constraint in (6.3) is then a normally distributed random variable.

As a result of this assumption, it is possible to rewrite the constraints without any random variables appearing in them. In the terminology of chance-constrained programming, the "deterministic equivalents" of the chance constraints can be found. This is possible because specifying the probability that a random variable with a known distribution be greater than or equal to some number is equivalent to specifying a fractile of the distribution. The fractiles of the normal distribution, however, can be written in terms of the mean and variance of the random variable. Hence, chance constraints of the type in (6.3) can be rewritten in terms of the decision variables -- the  $y_i$ 's and  $x_{ij}$ 's -- and the known means, variances, and covariances of the random variables -- the net cash outflows.

Assuming the net cash outflows are independently distributed normal random variables, the deterministic equivalent of the constraint in (6.3) would be

$$(6.4) \left\{ \begin{aligned} & \sum_{t=1}^s \sum_{i=1}^n E(c_{ti} - a_{ti})y_i + \sum_{t=1}^s \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n E(c_{tij} - a_{tij})x_{ij} - \sum_{t=1}^{s-1} r_L^t \ell + \sum_{t=1}^{s-1} r_B^t b_t \\ & + \ell_s - b_s + \left( \sum_{t=1}^s \sum_{i=1}^n \text{Var}(c_{ti} - a_{ti})y_i^2 + \sum_{t=1}^s \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n \text{Var}(c_{tij} - a_{tij})x_{ij}^2 \right)^{\frac{1}{2}} F^{-1}(\alpha_s) \\ & \cong \sum_{t=1}^s C_{t-1} \end{aligned} \right.$$

Notationally,  $E(z)$  denotes the expected value of  $z$ ,  $\text{Var}(z)$  denotes the variance

of  $z$ , and  $F(W)$  is the error function  $F(W) = \frac{1}{\sqrt{2\pi}} \int_0^W e^{-\frac{z^2}{2}} dz$  where  $z$  is a

standardized normal random variable. If some or all of the random variables had been correlated with one another, the square root on the left-hand side would have contained covariances of the random variables and cross products of the decision variables.<sup>14</sup>

In the deterministic-equivalent formulation, the decisionmaker's risk aversion appears as a financial transformation of the probability  $\alpha_s$ . The sum of all terms but the last on the left-hand side of (6.4) is the mean net cash outflow in period  $s$ . The right-hand side,  $\sum_{t=1}^s C_{t-1}$ , is the cumulated amount of corporate funds available in period  $s$ . The last term on the left-hand side of (6.4), the term to the one-half power multiplied by  $F^{-1}(\alpha_s)$ , represents the financial reserve the firm must hold above the mean net outflow if it is to obtain the safety margin it desires, as expressed by the probability  $\alpha_s$ . It is the reserve the firm must hold if it wants to ensure that its cumulated outflows do not exceed its cumulated inflows plus its cumulated internally generated funds more than  $(1-\alpha_s) \cdot 100$  per cent of the time, when each cumulation occurs through period  $s$ .

<sup>14</sup>As Naslund indicates, it is also not difficult to extend the model to the case where the  $C_{t-1}$ 's are normally distributed. Naslund (1966, p. 261).

Since a reserve may or may not be used in a given period, to ensure having the safety margin desired in period  $s$ , the flows in all periods up to and including  $s$  must be considered.

Having formulated the capital-budgeting problem as a problem of chance-constrained programming and transformed it into its deterministic equivalent, those following this course proceed in two directions. First, they use the Kuhn-Tucker conditions to discuss the acceptance and rejection criteria for projects. In particular, they focus on the effects of altering (i) the variance of a particular project, (ii) the risk level in a period, and (iii) the given riskless market rate of interest, on the acceptability of a project.<sup>15</sup> Second, they discuss the possibility of numerically solving the resulting problem, taking account of the integer restrictions on projects. The primary difficulty encountered in solving the problem so formulated is the presence of nonlinearities in the constraints, in the form of the radical in (6.4). Hillier and Naslund both introduce and use linear approximations for this square-root term.<sup>16</sup> The problem is thereby converted into a linear integer programming problem for which solution algorithms exist.

The analysis of the capital-budgeting problem using chance constraints is interesting. In particular, Naslund's discussion of the acceptability criteria for projects provides an interesting analogue to Weingartner's use of duality in the case of certainty models. But a fundamental shortcoming of the chance-constrained programming approach prevents the present author from viewing these models as anything but interesting theoretical exercises. The reason lies not with the lack of present operationalism of the approach for it certainly is no less practical than the present work's reliance on and emphasis on utility functions. Rather, the difficulty rests with the failure of the

---

<sup>15</sup>Naslund (1964, pp. 117-123, 126-131; and 1966, pp. 263-268).

<sup>16</sup>Hillier (1964, pp. 53-54) and Naslund (1966, pp. 268-271).

chance-constrained method to say anything about what happens if the values chosen for the decision variables do prove to be infeasible.

What happens the  $(1-\alpha_s) \cdot 100$  per cent of the time the sth budget constraint is violated? Alternatively, in terms of the deterministic equivalent constraint in (6.4), what happens if the financial reserves held fail to be adequate when a particular set of values of the random net outflows is realized? The reserve the chance-constrained approach leads the firm to hold will, after all, only provide that the sth cumulated budget constraint is met with a probability of  $\alpha_s$ , not with a probability of one. When such violations occur, the firm must have some contingency plan available. The failure of chance-constrained programming advocates to speak to this question leads the present writer to question seriously its usefulness as a decisionmaking tool in such risk situations.

#### 6.4.B. The Mean-Variance Approach: Its Rationale

A second approach to capital budgeting under risk that has been suggested by some theoreticians is the mean-variance approach. Several authors -- R. M. Adelson, J. Cord, and H. M. Weingartner -- have simply extended H. Markowitz's portfolio-selection model to the problem of a firm choosing investment projects.<sup>17</sup> In extending the approach, they do, of course, suitably modify the Markowitz model to take account of the discreteness of proposals, fixed budget dollars, physical interrelationships among projects, and other constraints subject to which the firm must make its decisions. J. Lintner, on the other hand, starts at a more fundamental level. He assumes that individual investors follow a mean-variance approach in reaching their individual portfolio decisions. Given certain further assumptions about the nature of the capital market, the expectations of individual investors about stock prices, and the expectations of individual investors and corporate managers concerning the

---

<sup>17</sup>Adelson (1965); Cord (1964); Weingartner (1966, pp. 498-504). The reference to Markowitz's model is to Markowitz (1959).

returns from potential corporate capital-investment projects, Lintner is then able to demonstrate that firms in such an environment should employ a mean-variance approach in budgeting their capital.<sup>18</sup>

No matter whether it is applied to the portfolio-selection problem of the individual investor or to the capital-budgeting problem of the firm, the mean-variance approach has the same basic rationale. It is grounded in the two fundamental assumptions Markowitz made about investors or at least about investors for whom his monograph Portfolio Selection was intended. Namely, "They want 'return' to be high," and "they want this return to be dependable, stable, not subject to uncertainty."<sup>19</sup> Return is then measured by the expected value of the payoff of the particular combination of securities or projects, while risk is measured by the variance of the portfolio's payoff. If two combinations of investments have the same expected return, the one whose payoff has a smaller variance is preferred; if two portfolios have the same variance of return, the one with the greater expected return is more desirable.

Markowitz presents a justification for his choice of measures of central tendency and variability.<sup>20</sup> In discussing the former, he compares the mean with the mode and the median, and he finds the mean the best of the three. For

---

<sup>18</sup>Lintner (1965). One of Lintner's basic assumptions is that individual investors and corporations face a perfect capital market. Any investor -- individual or corporate -- can borrow unlimited amounts at a riskless rate of interest, or can lend any amount, up to the limit of the individual's endowment or the firm's capital budget, at that same rate. When he discusses the capital-budgeting problem, he also assumes "that the investment opportunities available to the company in any time period are regarded as independent of the size and composition of the capital budget in any other time period." Lintner (1965, p. 28). These assumptions, among others, place Lintner's capital-budgeting model in a world very different from the one in which the model presented here is set.

<sup>19</sup>Markowitz (1959, p. 6).

<sup>20</sup>Ibid., pp. 49-55, 188-194, 286-297.

example, he shows that a small difference between two distributions can result in a large difference in their modes and that differences in mode may not accurately portray the direction or magnitude of differences in the distributions. On the other hand, the median of a probability distribution is insensitive to changes -- even if they are definitely good or definitely bad -- that do not alter the midpoint of the distribution. The mean, in contrast, is sensitive to all changes in the distribution, except those that exactly offset each other, and it always moves in the "correct" direction, accurately portraying the actual change in the distribution.

Turning to measures of variability of return, Markowitz considers five alternatives to the variance. They are the semi-variance, the expected value of loss, the expected absolute deviation of return, the probability of loss, and the maximum loss. For convenience, the alternatives and their definitions are summarized in (6.5), where  $Q$  is the stochastic payoff from the portfolio,  $\delta$  is an arbitrary threshold value chosen by the decisionmaker,<sup>21</sup> and  $E$  and  $Var$  have the same meaning as before.

$$(6.5) \left\{ \begin{array}{l} \text{Variance} = \text{Var } Q = E\{[Q-E(Q)]^2\} , \\ \text{Semi-variance}_\delta = E\{[(Q-\delta)^*]^2\} \\ \text{where } (Q-\delta)^* = \begin{cases} Q-\delta & \text{if } Q-\delta \leq 0 \\ 0 & \text{if } Q-\delta > 0 , \end{cases} \\ \text{Expected loss}_\delta = E\{-\min(Q-\delta, 0)\}, \\ \text{Expected absolute deviation}_\delta = E\{|Q-\delta|\} , \\ \text{Probability of loss} = E\{g(Q)\} \text{ where } g(Q) = \begin{cases} 0 & \text{if } Q > 0 \\ 1 & \text{if } Q \leq 0 , \end{cases} \\ \text{Maximum loss} = |\min Q| . \end{array} \right.$$

Markowitz's discussion of these alternative measures of dependability of return proceeds in an interesting way.

<sup>21</sup>The exact meaning of the parameter  $\delta$  depends on the particular measure one is discussing. Its meaning in the case of the semi-variance measure will become clearer shortly.

Taken as basic to his discussion is the assumption that the investor's behavior satisfies what has been alternatively called the expected-utility hypothesis, the expected-utility theorem, the Bernoulli principle, or (as Markowitz calls it) the expected-utility maxim. It states that the individual acts as if "(1) he attaches numbers, called their utility, to each possible outcome, and (2) when faced with chance alternatives he selects the one with the greatest expected value of utility."<sup>22</sup> The expected-utility maxim can be shown to be mathematically equivalent to what may be considered a more basic set of principles concerning the choices of a reasonable man in situations involving risk. This axiomatic approach to the expected-utility theorem was first presented by J. Von Neumann and O. Morgenstern in The Theory of Games and Economic Behavior.<sup>23</sup>

Using the axiomatic approach to the expected-utility maxim, Markowitz shows<sup>24</sup> that an individual (i) maximizes the expected value of some utility function and (ii) bases his preferences solely on the expected payoff,  $E(Q)$ , and some measure of the risk of that payoff,  $H(Q)$ , where  $H(Q)$  is itself the expected value of some function of the payoff  $Q$ , say  $H(Q) = E\{h(Q)\}$ , if and only if the individual maximizes the expected value of a utility function  $u(Q) = aQ + bh(Q)$ . Since Von Neumann-Morgenstern utility functions are unique only up to a positive linear transformation, it is legitimate to choose the zero of the utility scale so that  $u(0) = 0$ . This choice is reflected in the absence of a constant term from the utility function  $u(Q) = aQ + bh(Q)$ . With the aid of this result the alternative measures of risk can be evaluated.

First, Markowitz proves that if one assumes the decisionmaker maximizes the expected value of a utility function,  $u(Q)$ , then it is not possible that the

---

<sup>22</sup>Markowitz (1959, p. 208).

<sup>23</sup>See Von Neumann and Morgenstern (1947, Chapter 1, Section 3, especially pp. 26-29, and Appendix). For two other good discussions of the axiomatic approach to the expected-utility maxim, see Luce and Raiffa (1957, Chapter 2, especially pp. 23-32) and Markowitz (1959, Chapter X).

<sup>24</sup>Markowitz (1959, pp. 286-287).

decisionmaker acts on the basis of mean and maximum loss. There is a contradiction between the use of maximum loss and the Von Neumann-Morgenstern utility axioms. The use of any of the other measures of risk in conjunction with a portfolio's expected return, however, is consistent with the decisionmaker's maximizing the expected value of a single utility function  $u(Q)$ .

Examining the utility functions associated with expected loss, expected absolute deviation, and probability of loss, Markowitz concludes that these measures do not provide trustworthy evaluations of portfolios. The resulting combinations of investments "can be foolishly speculative even when apparently conservative."<sup>25</sup> The shortcomings of these three measures rest with the presence of certain undesirable linear segments and/or discontinuities in their associated utility functions.

The utility functions associated with each of these three measures have long straight-line segments in the range of negative returns,  $Q < 0$ . This means the investor using any one of these measures of risk would be indifferent to at least some actuarially neutral risks whose outcomes were completely in the range of negative payoffs. He would never buy insurance if the insurance company had the smallest expected gain and he would be exactly indifferent to buying insurance if the company's expected gain were zero. Markowitz rejects all three of these measures of risk because their utility functions' linear segments in the negative-returns region show their true lack of conservatism. They do not accurately warn the expected-utility maximizer about a risky situation. The inappropriateness of the probability-of-loss measure is compounded by the discontinuous nature of its associated utility function. Since the presence of a discontinuity means the function cannot be concave everywhere, there are some actuarially neutral risks that the individual using this measure would prefer

---

<sup>25</sup>Ibid., p. 297. Markowitz's demonstration of and evaluation of the utility functions associated with the alternative measures of risk appears in Markowitz (1959, pp. 287-297).



to some corresponding status quo positions. In short, there are some gambles for which the individual basing his decision upon expected return and probability of loss would actually be a risk-lover!

This leaves Markowitz with the variance and semi-variance as alternative measures of risk. In comparing the advantages and disadvantages of these two measures, he notes that "Variance is superior with respect to cost, convenience, and familiarity." For example, "In an analysis based on V [variance], only means, variances, and covariances must be supplied as inputs; whereas an analysis based on S [semi-variance] requires the entire joint distribution of returns." But, he goes on, "the requirement, by analyses based on variance, for covariances between each pair of securities leads to models for derived covariances -- models which are sometimes equivalent to, sometimes little less than, models of the entire joint distribution required by an analysis based on S."<sup>26</sup> With regard to the questions of computation and familiarity, Markowitz also indicates that the variance measure's relative advantages are not all that great. He writes, "the computing cost involved even in the use of semi-variance is small compared with the other research costs involved in supervising one or more large portfolios [or capital budgets!]. . . . Familiarity, finally, is a transient thing: use can make S as familiar as V."<sup>27</sup>

On the other hand, use of the semi-variance instead of the variance seems to yield considerable benefits, as Markowitz himself notes. While both consider very low returns undesirable, the variance has the added peculiarity of considering very high returns equally undesirable. A glance at their definitions in (6.5) reveals how this results. Suppose  $\delta$  in the definition of semi-variance is set equal to the mean of  $Q$ , that is,  $\delta = E(Q)$ . Next note that since variance and semi-variance are meant to measure risk, which is undesirable from the investor's point of view, the higher they are, the worse the portfolio supposedly is.

---

<sup>26</sup>Ibid., pp. 193-194.

<sup>27</sup>Ibid., p. 194.

Consider, first, a very low return that occurs with a given probability. The possibility that this return occurs is clearly an undesirable feature of the portfolio. From (6.5), with  $\delta = E(Q)$ , it is seen that this particular return's deviation from the portfolio's average return enters into the calculation of both the variance and the semi-variance. In fact, this undesirable deviation from the mean contributes equally to the variance and the semi-variance. Now consider a desirable deviation from the mean, a very high return that occurs with a given probability. For this possible realization of  $Q$ ,  $[Q-E(Q)]^2$  is positive while  $[Q-E(Q)]^+$  is zero. Hence, this desirable possibility of a very high return increases the variance of the portfolio making the mix of investments look worse, but it does not affect the semi-variance. In short, an analysis based on the semi-variance about the mean concentrates on reducing losses, while an analysis using the variance concentrates on eliminating extreme returns in either direction.

Alternatively stated, the semi-variance takes account not only of the variability of the investment mix's return but also of the skewness of the return's probability distribution. If the returns of two portfolios, say A and B, show an equal amount of variability from the mean, the variance measure will indicate that they have the same degree of risk. If, however, portfolio A shows greater extreme returns to the right than to the left as compared with portfolio B, the semi-variance will show that A is less risky than B. As Markowitz notes, the ratio of the variance to twice the semi-variance about the mean ( $S_E$ ), that is,  $\frac{\text{Var}}{2S_E}$ , can be used as a measure of skewness. A symmetric distribution will have  $\frac{\text{Var}}{2S_E}$  equal to unity, a distribution skewed right will have this ratio greater than unity, one with extremes to the left outweighing those to the right will have the ratio less than unity.<sup>28</sup>

---

<sup>28</sup> Ibid., pp. 190-191.

In Markowitz's terminology, a portfolio is "efficient" if it simultaneously maximizes expected return for a given level of risk and minimizes the degree of risk for a given level of expected return. If a portfolio is efficient, one cannot obtain a higher average return without accepting greater variability of return and one cannot obtain a smaller variability (greater certainty of return) without sacrificing some average return.<sup>29</sup> What, then, can one say about the efficient portfolios generated by an analysis based on variance as compared to those resulting from an analysis using semi-variance? First, if all distributions were symmetric or had the same degree of skewness, the two measures when used with expected return would yield the same set of efficient investment mixes. Second, if the portfolios' returns had probability distributions with different degrees of skewness, among portfolios with the same mean and variance, an analysis based on  $S_E$  would choose the portfolio with the greatest rightward skewness or least leftward skewness.

Markowitz concludes, "There is always a portfolio efficient in terms of expected return and  $S_b$  (for suitable  $b$ ) which is at least as good as the best portfolio based on analysis using expected return and variance."<sup>30</sup> To put his use of variance in a more favorable light, however, he also writes that "Efficient portfolios based on variance, however, cannot be characterized as bad or undesirable.... The only complaint one can raise about such a portfolio is that it sacrifices too much expected return in eliminating both extremes."<sup>31</sup> As shall be seen presently, this is, in fact, not the only complaint one can raise against an analysis based on mean and variance. Interestingly enough, remedying one of the major problems with the mean-variance approach will involve introducing a measure of skewness (although one different from  $\frac{\text{Var}}{2S_E}$ ) into the evaluation of investment portfolios.

---

<sup>29</sup>Ibid., pp. 6, 129.

<sup>30</sup>Ibid., p. 297.

<sup>31</sup>Ibid., p. 194.

6.4.C. The Mean-Variance Approach: Its Versions and Origins

If, however, one does accept the expected value of the payoff of the investment program as the measure of return and the variance of that payoff as the measure of undependability of return, several alternative formulations of the problem of capital budgeting under risk can be found in the literature. Omitting, for the present, the constraints deriving from fixed budget ceilings, borrowing limits, physical interrelationships, other limited resources, and project discreteness, in order to focus on the mean-variance objective, three versions of the approach can be identified. The first is

$$(6.6) \quad \left\{ \begin{array}{l} \text{Maximize } E(\text{Payoff}) \\ \text{Subject to } \text{Var}(\text{Payoff}) \leq V_0 \end{array} \right.$$

where  $V_0$  is a preassigned tolerance level for variance. The second is

$$(6.7) \quad \left\{ \begin{array}{l} \text{Minimize } \text{Var}(\text{Payoff}) \\ \text{Subject to } E(\text{Payoff}) \geq E_0 \end{array} \right.$$

where  $E_0$  is a preassigned level of expected return. Lastly, the mean-variance model has appeared in the form

$$(6.8) \quad \text{Maximize } E(\text{Payoff}) - \lambda \text{Var}(\text{Payoff}) ,$$

subject to the constraints the firm faces.<sup>32</sup> The parameter  $\lambda$  in the objective function in (6.8) is interpreted as a measure of risk aversion. It indicates the decisionmaker's rate of substitution between increases in the expected return

---

<sup>32</sup>The model in (6.6) appears in Cord (1964), the one in (6.7) is used by Adelson (1965) to trace out the "efficient set" for the capital-budgeting problem, and the version in (6.8) is presented in Lintner (1965) and Weingartner (1966). This classification of the several versions of the mean-variance approach is also to be found in Hakansson (1966, pp. 111-112).

of the investment program and increases in the variance of the program's payoff.<sup>33</sup>

If one accepts the expected-utility theorem as the fundamental basis for behavior under risk, as the present study does, there are two and only two possible derivations of the mean-variance approach.<sup>34</sup> A risk-averting investor with a Von Neumann-Morgenstern single-period utility function  $u(Q)$  will make his decisions on the basis of mean and variance alone if and only if (i) his utility function is quadratic in  $Q$  or (ii) his utility function is concave and the returns on the individual investments have a multivariate distribution with the property that the distribution of a linear combination of these returns is completely specified by its first two moments, or both (i) and (ii) are true. These results are due to Borch and Tobin, and they can be illustrated briefly.<sup>35</sup>

Notationally, let  $a_i$  denote the single-stage stochastic payoff from investment  $i$ . This  $i$ th investment is undertaken to the extent indicated by the decision variable  $y_i$ . If there are  $N$  possible investments, the payoff of the portfolio of investments is the random variable  $Q = \sum_{i=1}^N a_i y_i$ . The decisionmaker then maximizes

$$(6.9) \quad E\{u(Q)\} = \int_{-\infty}^{\infty} u(Q) dF(Q) ,$$

where  $F(Q)$  is the probability distribution of the payoff of the investment program.

<sup>33</sup>Note that what a model of the type in (6.8) is attempting to do is determine a certainty equivalent for the stochastic payoff from the investment program. It tries to reduce the return and risk dimensions to a single figure from which risk has been removed. The risk-discount approach presented earlier is another way of trying to determine a certainty equivalent. Another popular method for collapsing risk and return into a single figure that some writers refer to as the certainty-equivalent approach simply multiplies the expected return of the portfolio by a factor that summarizes the effects of risk on value. The resulting product is to be such that the decisionmaker is indifferent between receiving it and facing the stochastic payoff. The multiplying factor will be larger, these writers say, the smaller the dispersion of the distribution of returns, the more attractive the form of the distribution is to the investor, and so on. See, for example, Robichek and Myers (1965, pp. 80-86).

<sup>34</sup>See Roy (1952) for the derivation of a decisionmaking objective function involving only the mean and the variance of the payoff that is not based on a Von Neumann-Morgenstern utility function.

<sup>35</sup>Borch (1963) and Tobin (1958, pp. 74-77).

Consider, first, the case in which  $F(Q)$  is an arbitrary probability distribution. The objective function in (6.9) clearly can then be a function of  $E(Q)$  and  $\text{Var}(Q)$  alone if and only if  $u(Q)$  is a second-degree polynomial. Setting  $u(0) = 0$ , as one may do since the utility function is unique only up to a positive linear transformation, it can be assumed that

$$(6.10) \quad u(Q) = \beta Q - \gamma Q^2 \quad \text{with } \gamma > 0.$$

The sign of  $\gamma$  is dictated by the fact that the decisionmaker is risk-averse, for risk aversion is equivalent to concavity of the utility function or  $u''(Q) < 0$ . Substituting (6.10) into the general objective function in (6.9) one obtains

$$E\{u(Q)\} = \int_{-\infty}^{\infty} [\beta Q - \gamma Q^2] dF(Q) = \beta E(Q) - \gamma E(Q^2) \quad \text{or}$$

$$(6.11) \quad E\{u(Q)\} = \beta E(Q) - \gamma [E(Q)]^2 - \gamma \text{Var}(Q).$$

This is, in fact, the only form an objective function in mean and variance alone can assume if it is to be consistent with an underlying Von Neumann-Morgenstern utility function, and no restrictions are placed on the distribution function  $F(Q)$ . In particular, without a restriction on that distribution function, the objective function in (6.8) is not a legitimate one to use if the expected-utility theorem is accepted. Also, accepting that maxim, it is incorrect to use an objective function of the form

$$(6.12) \quad \text{Maximize } \beta E(Q) - \gamma \sqrt{\text{Var}(Q)},$$

that is, a linear function of mean and standard deviation alone.<sup>36</sup>

Moreover, while the objective function in (6.8) can be legitimate if the payoffs have a particular multivariate distribution, it does not seem

---

<sup>36</sup>This observation is also due to Borch. See Borch (1965, Chapter IV, pp. 25-26).

possible to rescue the objective function in (6.12) by restricting consideration to particular distributions of returns. The author cannot offer a general proof of this conjecture, but the statement can easily be seen to be true for any utility function for which a Taylor's series representation exists.<sup>37</sup> This is unfortunate because an objective function of the form in (6.12), linear in mean and standard deviation, has appeared in several different places in the literature.<sup>38</sup>

Turning to the second possible way that a mean-variance approach can be derived, suppose the distribution of the portfolio's payoff is completely determined by its first two moments. Then if  $u(Q)$  is a function for which the integral in (6.9) converges, it is clear that one will have

$$(6.13) \quad E\{u(Q)\} = \int_{-\infty}^{\infty} u(Q) dF(Q) = g(E(Q), \text{Var}(Q)) ,$$

---

<sup>37</sup>For a statement of conditions under which such a representation exists, see J. M. H. Olmsted (1961, pp. 418-420).

<sup>38</sup>Van Moeseke's "truncated minimax" criterion, Van Moeseke (1965), is precisely of the form in (6.12). He assumes that (in our notation)  $Q$  is normally distributed. But he also asserts that his criterion has a reasonable interpretation in terms of weighting mean and standard deviation even in the absence of the normality assumption. Van Moeseke (1965, p. 209). If  $Q$  is not assumed to be a random variable with a two-parameter distribution, the objective function in (6.12) cannot possibly result if one accepts the expected-utility theorem. Assuming  $Q$  is, say, normally distributed does not solve the problem. One then has to seek a  $u(Q)$  without a Taylor's series representation that will produce (6.12) from the integration in (6.9). It is not at all clear that this is possible. Attempts on the present author's part to find such a function proved fruitless.

For other uses of an objective function of the type in (6.12), see Farrar's discussion of previously proposed certainty equivalents. Farrar (1962, p. 15). Note also that if one accepts Baumol's modification of the Markowitz model, using  $E(Q)$  and  $E(Q) - K\sqrt{\text{Var}(Q)}$  (where  $K$  is a parameter) instead of  $E(Q)$  and  $\text{Var}(Q)$ , for the analysis, problems might also arise. Specifically, the use of linear indifference curves -- although not suggested by Baumol -- would lead to an objective function of the type in (6.12) and its associated problems. For a discussion of Baumol's suggestion, see Baumol (1963) and Russell and Smith (1966).

where  $g$  is a function whose form will depend on  $u(Q)$  and  $dF(Q)$ . For example, R. J. Freund<sup>39</sup> takes the utility function (in our notation) to be

$$(6.14) \quad u(Q) = 1 - e^{-2\lambda Q} ,$$

where  $\lambda$  indicates the degree of risk aversion -- the higher  $\lambda$  is, the more risk-averse is the decisionmaker. Assuming  $Q$  to be normally distributed, he shows that maximizing the expected value of that utility function is equivalent to maximizing the objective function in (6.8), namely

$$(6.8) \quad E\{u(Q)\} = E(Q) - \lambda \text{Var}(Q) .$$

Since  $Q = \sum_{i=1}^N a_i y_i$ , it is clear that if the distribution of  $Q$  is to be governed by two parameters, the returns of the individual investments must have a multivariate distribution such that a linear combination of these returns has a distribution that is so determined. As Pye has written recently, "The multivariate normal is an important example of such a multivariate probability distribution. However, it may not be easy [if it is possible at all!] to find others when the returns are not independent."<sup>40</sup> And, of course, stochastic interdependence among investment returns is an essential aspect of the portfolio problem of the individual investor and the capital-budgeting problem of the firm.

#### 6.4.D. The Mean-Variance Approach: Some Objections

Both of these origins of the mean-variance approach, the quadratic utility function and the assumption of a multivariate normal distribution for the returns from investments, are open to strong and convincing objections. Since these are the only two possible derivations of the mean-variance approach once the expected-utility hypothesis is accepted, the presence of severe shortcomings in these

<sup>39</sup> Freund (1956, pp. 254-255).

<sup>40</sup> Pye (1967, p. 111).



sources seriously weakens the underpinnings of the mean-variance approach itself.

The first objection put forth against the quadratic utility function in (6.10) is its implication that positive returns have negative marginal utility over a certain range.<sup>41</sup> Specifically, the utility function in (6.10) reaches a maximum when  $Q = \frac{\beta}{2\gamma}$ . For returns smaller than  $\frac{\beta}{2\gamma}$ , added returns increase the decisionmaker's utility; but for  $Q > \frac{\beta}{2\gamma}$ , increasing the portfolio's return decreases his utility. This phenomenon results from the need to have the coefficient of the squared term,  $-\gamma$  in this case, negative if the function is to be concave and hence belong to a risk-averter. This objection can be met, however, by a proper choice of parameters for the quadratic. One could then ensure that  $\frac{\beta}{2\gamma}$  is large enough so that the utility function shows positive marginal utility for the entire range of relevant returns.

A second objection to the quadratic cannot be met so easily. The function has an implausible property that no alteration of the parameters could rectify. Namely, it shows increasing absolute risk aversion throughout the entire range of returns. As the individual decisionmaker becomes wealthier, he actually decreases his total holdings of risky assets if he possesses a quadratic utility function. Arrow puts it quite strongly when he writes, as income increases, "the willingness to gamble for a bet of fixed size will necessarily decrease, a result which shows the absurdity of the quadratic assumption."<sup>42</sup> If investors had quadratic utility functions, risky investments would be inferior goods in that "Offering more return at the same risk would so sate investors that they would reduce their risk-investments because they were more attractive."<sup>43</sup>

---

<sup>41</sup>This objection has been discussed in Borch (1963), Hirshleifer (1965, p. 521), and Lintner (1965, p. 18).

<sup>42</sup>Arrow (1965, p. 36).

<sup>43</sup>Lintner (1965, p. 18).

This property of the quadratic utility function was first noted by Arrow and Pratt.<sup>44</sup> That the quadratic does, in fact, exhibit increasing absolute risk aversion everywhere can be seen quite easily. Recall from Section 2.2 that a single-stage utility function shows decreasing, constant, or increasing absolute risk aversion as  $-\frac{u''(W)}{u'(W)}$  decreases, is constant, or increases with increases in  $W$ . For increasing (more precisely, nondecreasing) absolute risk aversion one must therefore have

$$(6.15) \quad u'(W)u'''(W) \cong [u''(W)]^2 .$$

But for the quadratic in (6.10), replacing  $Q$  with  $W$ , one has  $u'(W) = \beta - 2\gamma W$ ,  $u''(W) = -2\gamma$ , and  $u'''(W) = 0$ . Hence, the condition in (6.15) is fulfilled since  $4\gamma^2 > 0$ .

If the mean-variance approach had to rest the case for its use upon the quadratic utility function, its usefulness therefore would be highly questionable. The undesirable and implausible property of increasing risk aversion which marks the quadratic utility function makes it and the mean-variance approach based upon it an inappropriate objective function for a model of capital budgeting under risk. Unfortunately, the second possible source of the mean-variance approach does not offer much in the way of plausibility either. It, too, seems to be a rather weak foundation upon which to base an objective function for such a model. In discussing the individual investor's portfolio problem Lintner writes, "despite its mathematical convenience, multivariate normality is doubtless also suspect, especially perhaps in considering common stocks."<sup>45</sup> I would append Lintner's comment with the phrase "and capital-investment projects."

It is difficult to believe that the joint distribution of gross returns on projects can be adequately represented by a multivariate normal distribution

---

<sup>44</sup>Arrow (1963, p. 26); Pratt (1964, p. 132).

<sup>45</sup>Lintner (1965, pp. 18-19).

when interdependence among projects is to be recognized. It is hard to find plausible the assumption that none of the subjective distributions of gross returns will show any skewness, that all will have a similar smooth bell-shaped distribution. Certainly, not all projects will be supposed to have symmetric chances of doing well and doing poorly. If, though, one restricts oneself to two-parameter families of distributions, and normal distributions in particular, indications of such skewness are lost. It seems that too much of reality is sacrificed by such a restrictive assumption.

This catalogue of objections has pertained to the single-stage utility function  $u(Q)$  or  $u(W)$ . When the mean-variance approach has been applied to capital budgeting under risk it has been assumed that each project's stream of inflows and outflows could somehow be reduced to a single figure. Cord takes as a project's figure of merit its internal rate of return, Adelson and Lintner use its discounted present value calculated at the riskless rate that exists in each of their models, and Weingartner uses the generic single figure of merit "payoff." The present study, in contrast, has placed a multiperiod utility function at the center of the capital-budgeting problem. Accepting the expected-utility theorem, the decisionmaking unit in the present model maximizes

$$(6.16) \quad E\{U(W_1, \dots, W_T)\} = \int \dots \int_A U(W_1, \dots, W_T) f(W_1, \dots, W_T) dW_1 \dots dW_T,$$

where  $f(W_1, \dots, W_T)$  is the multivariate frequency function of the  $T$  stochastic consumption alternatives and  $A$  is the region of integration defined by the ranges of the  $W_t$ 's. Each  $W_t$  is defined by equation (5.15) to be

$$(6.17) \quad W_t = \sum_{i=1}^n a_{ti} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n a_{tij} x_{ij} + (1+r_L)R_t.$$

The random elements of  $W_t$ , it will be recalled, are the returns to the individual real projects, the  $a_{ti}$ 's, and the cash-flow interactions among the projects, the

$a_{tij}$ 's. The  $y_i$ 's,  $x_{ij}$ 's, and  $R_t$ 's are decision variables, and the payoff of the cash-withdrawal activity, the level of which is indicated by  $R_t$ , is nonstochastic.

In this framework, there is no longer a single figure that can be called the payoff of the investment program since the capital budget's payoff is now the stream of returns  $W_1, W_2, \dots, W_T$ . Without such a figure, one cannot speak of a mean-variance approach as one that evaluates alternative programs on the basis of two figures: an average return and the variance of the return. But an approach comparable to the single-period mean-variance approach is possible when the decision must be made in a multiperiod setting.

The essence of the mean-variance approach, after all, is that only the first two moments of all the relevant probability distributions need to be considered. The investor -- whether an individual or a corporation -- can choose the optimal investment plan knowing only the mean and variance of each investment's return and the covariance of each return with every other one. In a multiperiod context, then, one can refer to as a mean-variance approach a procedure that enables investment programs to be chosen on the basis of an analogous set of data. Namely, a mean-variance procedure would be one that required as data only (although it still is quite a bit of information to ask for) the means, variances, and covariances of the random variables determining the stream of stochastic withdrawals -- that is, the mean and variance of each  $a_{ti}$  and each  $a_{tij}$  and the covariance between every pair of  $a_{ti}$ 's, every pair of  $a_{tij}$ 's, and every  $a_{ti}, a_{tij}$  combination.

The questions that arise are: Under what conditions will such an approach be possible and do these origins have the same shortcomings as the two alternatives in the single-stage mean-variance case? Once again, given the expected-utility theorem as the basis for behavior under risk, there are two and only two cases in which a risk-averting decisionmaker would act on the basis of means, variances, and covariances alone. If and only if (i) the utility function were

quadratic in the  $W_t$ 's or (ii) the utility function were concave and the multivariate frequency function  $f(W_1, W_2, \dots, W_T)$  completely specified by the first and second moments (individual and joint) of the individual-project returns and cash-flow interactions, or (i) and (ii) both were true, would maximizing the objective function in (6.17) result in a mean-variance approach. Unfortunately, as in the single-stage case, both of these possible origins of a multiperiod mean-variance approach are open to serious charges of implausibility.

Using conditions on the probability density function  $f(W_1, W_2, \dots, W_T)$  to arrive at the mean-variance result entails assumptions even more questionable than those involved in the single-period case. One must again assume that the distributions of the  $a_{ti}$ 's and  $a_{tij}$ 's are members of a two-parameter family of distributions such that a linear combination of these variables for a given  $t$  also has a two-parameter distribution. The multivariate normal is, as was suggested earlier, the only common distribution known to the author which has this property. But in the multiperiod case an even stronger supposition is required.

The multivariate density function of the resulting linear combinations, the  $W_t$ 's, must also be taken to be completely determined by the means, variances, and covariances of the  $W_t$ 's. If the  $a_{ti}$ 's and  $a_{tij}$ 's for a given  $t$  have a multivariate normal distribution, the necessary requirement on  $f(W_1, W_2, \dots, W_T)$  can be fulfilled if it is assumed the latter is also a multivariate normal or that the  $W_t$ 's are independent. It should be noted carefully that one of these assumptions must be made explicitly. It does not suffice to have each of the  $W_t$ 's normally distributed because the marginal densities of a joint density function can be normal and yet the joint density function be non-normal. In fact, the marginal density functions can be normal while the joint density function does not even exist.<sup>46</sup> Hence, it seems that the course one must follow if a multiperiod

---

<sup>46</sup>For a discussion and examples see Anderson (1958, pp. 37-38) and Feller (1966, pp. 84, 99).

mean-variance approach is to be derived from assumptions about the underlying probability distributions is to assume that the individual returns in all periods and the cash-inflow interactions in all periods are governed by a joint multivariate normal density function. The objections to multivariate normality when it was assumed about single-payoff measures clearly hold a fortiori when it is assumed about all relevant cash inflows in all periods.

On the other hand, the assumption that the utility function  $U(W_1, W_2, \dots, W_T)$  is quadratic in its arguments is open to the same objections as in the case of the single-stage quadratic utility function. First, since as was proven in Chapter 2,  $U(W_1, W_2, \dots, W_T)$  must be concave if the decisionmaker is to be risk-averse, the coefficients of each of the  $W_t^2$  terms must be negative.<sup>47</sup> That is, if the general multiperiod quadratic utility function is written as

$$(6.18) \quad U(W_1, W_2, \dots, W_T) = \sum_{t=1}^T (\beta_t W_t - \gamma_{tt} W_t^2) + \sum_{s=1}^T \sum_{\substack{t=1 \\ s \neq t}}^T \gamma_{st} W_s W_t,$$

then each  $\gamma_{tt}$  must be positive. But, then, analogous to the single-stage case, there will exist ranges of each period's income over which increasing the program's returns in that particular period will actually decrease the decisionmaker's utility. There will be regions of the T-dimensional consumption-incomes space in which consumption income in one or more periods will have negative marginal utility.

In particular, consumption income in period t will have negative marginal utility whenever

---

<sup>47</sup>It will be recalled that a twice-differentiable function is concave (convex) if and only if its second total differential is a negative (positive) definite quadratic form. A quadratic form is negative definite if and only if the naturally ordered principal minors of the form's Hessian alternate in sign, beginning with negative for the first-order principal minor. A quadratic form is positive definite if and only if all the naturally ordered principal minors of the form's Hessian are positive. See Hadley (1961, pp. 259-262) for a discussion of these determinantal conditions.

$$(6.19) \quad W_t > \frac{\beta_t + \sum_{\substack{s=1 \\ s \neq t}}^T \gamma_{st} W_s}{2\gamma_{tt}}$$

Note carefully that the particular values of  $W_t$  for which added income in the  $t$ th period actually reduces utility depend on the values assumed by the consumption incomes in the other periods, as indicated by the sum in the numerator. This is the reason for referring to "regions" of the consumption-incomes space in which different periods' incomes will have negative marginal utility. The ranges of  $W_t$ -values over which added consumption income in the  $t$ th period decreases total utility will differ for each possible configuration of the other  $T-1$  consumption incomes. Although the task is now more complicated, the problem of negative marginal utilities can be circumvented just as it was in the single-stage case. Specifically, by a proper choice of parameters,  $\beta_t$ 's,  $\gamma_{tt}$ 's, and  $\gamma_{st}$ 's, one can ensure that the utility function in (6.18) shows positive marginal utility for every period's consumption income for all relevant consumption-income vectors,  $W = (W_1, W_2, \dots, W_T)$ .

Even if this objection is met, however, the multiperiod quadratic utility function in (6.18) is unsuitable for use when a risk-averse decisionmaker is choosing a course of action in a risk environment. The utility function in question has the property of being increasingly risk-averse no matter what values the parameters assume. To prove this it will be necessary to recall some notation and definitions from Chapter 2.

The vector  $\tilde{Z}$  denotes the vector of actuarially neutral risks the decisionmaker faces,  $\tilde{Z} = (\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_T)$ , with one risk appearing in each period.<sup>48</sup> The vector  $\Pi = (\pi_1, \pi_2, \dots, \pi_T)$  represents a vector of insurance premiums or risk

---

<sup>48</sup> Recall that actuarially neutral means  $E(\tilde{Z}) = 0$ , that is,  $E(\tilde{z}_t) = 0$  for all  $t$ .

premiums the decisionmaker would be willing to pay in order to avoid facing the given risks. That is, if his asset or income vector were  $W = (W_1, W_2, \dots, W_T)$ , then he would be exactly indifferent to receiving  $W - \Pi$ , where  $\Pi$  depends on  $W$  and  $\tilde{Z}$ , or facing the given vector of risks. An insurance policy  $\Pi = \Pi(W, \tilde{Z})$  is thus a vector of money amounts such that

$$(6.20) \quad U(W - \Pi) = E\{U(W + \tilde{Z})\},$$

because the decisionmaker is taken to accept the expected-utility theorem as the basis for his behavior under risk. In particular,  $\Pi^t = (0, \dots, 0, \pi_t^t, 0, \dots, 0)$  is an insurance policy in which the decisionmaker makes a payment in only the tth period when he pays the amount  $\pi_t^t$ . That is,  $\Pi^t$  is a vector with zeros everywhere but in the tth element.

As the reader will recall, it was shown that for a strictly risk-averse decisionmaker  $\pi_t^t$  is positive for all  $t$ . That is, for a strict risk averter all  $T$  of the  $\Pi^t$ -vectors are semipositive. Last, and of great importance to the present question, a decisionmaker was defined to be decreasingly risk-averse if as  $W$  increases -- that is, as at least one  $W_t$  increases and none decreases --  $\Pi^t(W, \tilde{Z})$  is nonincreasing for all  $t$  and strictly decreasing for at least one  $t$  for all risks  $\tilde{Z}$ . Similarly, a decisionmaker would be increasingly risk-averse if as  $W$  increases,  $\Pi^t(W, \tilde{Z})$  were nondecreasing for all  $t$  and strictly increasing for at least one  $t$  for all risks  $\tilde{Z}$ .<sup>49</sup>

With this notation and set of definitions in hand, the following theorem can be proven.

---

<sup>49</sup>Recall that whereas in the case of a single-stage utility function, constant, decreasing, and increasing risk aversion are exhaustive possibilities, in the case of a multiperiod function they are not. While still mutually exclusive, our alternatives are not mutually exhaustive in the multiperiod case. See the discussion in Section 2.5.A.



THEOREM 6.1. If  $U(W)$  is quadratic in  $W = (W_1, W_2, \dots, W_T)$  and the decisionmaker is a risk averter, then the decisionmaker cannot be decreasingly risk-averse. In fact, he will be increasingly risk-averse.

PROOF: Suppose the initial consumption-income vector is  $\bar{W} = (\bar{W}_1, \bar{W}_2, \dots, \bar{W}_T)$ . From the definition of an insurance-premium vector, it follows that at the initial position

$$(6.21) \quad U(\bar{W} - \Pi^t(\bar{W}, \tilde{Z})) = E\{U(\bar{W} + \tilde{Z})\} \quad \text{for all } t.$$

Now suppose the consumption-income vector increases by a set of incremental changes  $d\bar{W}_i \geq 0$  for all  $i$  and  $d\bar{W}_i > 0$  for at least one  $i$ . Since the condition in (6.21) must hold at both the initial and final positions, that is, at  $\bar{W}$  and at  $\bar{W} + d\bar{W}$ , for small enough  $d\bar{W}$  one has

$$(6.22) \quad d[U(\bar{W} - \Pi^t(\bar{W}, \tilde{Z}))] = d[E\{U(\bar{W} + \tilde{Z})\}] \quad \text{for all } t,$$

where the differential is with respect to the  $\bar{W}_i$ 's. Since the differential and expectations operations are commutative, the equation in (6.22) may be rewritten as

$$(6.23) \quad d[U(\bar{W} - \Pi^t(\bar{W}, \tilde{Z}))] = E\{dU(\bar{W} + \tilde{Z})\}.$$

But

$$(6.24) \quad d[U(\bar{W} - \Pi^t(\bar{W}, \tilde{Z}))] = \sum_{i=1}^T \left( \frac{\partial U}{\partial \bar{W}_i} \right)_{\bar{W} - \Pi^t} d\bar{W}_i - \sum_{i=1}^T \left( \frac{\partial U}{\partial \bar{W}_t} \right)_{\bar{W} - \Pi^t} \frac{\partial \pi_t^t}{\partial \bar{W}_i} d\bar{W}_i,$$

where a subscript on a partial derivative (or on a differential in what follows) indicates the point at which it is to be evaluated and  $\Pi^t(\bar{W}, \tilde{Z})$  is written as  $\Pi^t$  where no confusion is possible. Recall also that  $\pi_t^t$  is the tth and only nonzero element of  $\Pi^t$ . But  $\left( \frac{\partial U}{\partial \bar{W}_t} \right)_{\bar{W} - \Pi^t}$  is clearly independent of  $i$  and since only

the  $\bar{W}_i$ -elements are changing,  $d\pi_t^t = \sum_{i=1}^T \frac{\partial \pi_t^t}{\partial \bar{W}_i} d\bar{W}_i$ . Hence, (6.24) is

equivalent to

$$(6.25) \quad d[U(\bar{W} - \Pi^t(\bar{W}, \tilde{Z}))] = [dU(W)]_{\bar{W} - \Pi^t} - \left( \frac{\partial U}{\partial W_t} \right)_{\bar{W} - \Pi^t} d\pi_t^t,$$

since  $\sum_{i=1}^T \left( \frac{\partial U}{\partial \bar{W}_i} \right)_{\bar{W} - \Pi^t} d\bar{W}_i = [dU(W)]_{\bar{W} - \Pi^t}$ .

Turning to the right-hand side of (6.23), since  $U(W)$  is quadratic,  $dU(\bar{W} + \tilde{Z})$  is a linear form in the  $\bar{W}_t + \tilde{Z}_t$  arguments. Therefore

$$(6.26) \quad E\{dU(\bar{W} + \tilde{Z})\} = dU(E\{\bar{W} + \tilde{Z}\}) = dU(\bar{W}) = [dU(W)]_{\bar{W}}$$

since  $E\{\tilde{Z}\} = 0$  and  $\bar{W}$  is nonstochastic.

Combining the results in (6.25) and (6.26), one obtains

$$(6.27) \quad [dU(W)]_{\bar{W} - \Pi^t} - \left( \frac{\partial U}{\partial W_t} \right)_{\bar{W} - \Pi^t} d\pi_t^t = [dU(W)]_{\bar{W}}$$

and

$$(6.28) \quad d\pi_t^t = \frac{1}{\left( \frac{\partial U}{\partial W_t} \right)_{\bar{W} - \Pi^t}} \{ [dU(W)]_{\bar{W} - \Pi^t} - [dU(W)]_{\bar{W}} \} \text{ for each } t.$$

But since it is assumed that the first objection to the quadratic utility function is met, it is known that  $\left( \frac{\partial U}{\partial W_t} \right)_{\bar{W} - \Pi^t}$ , the marginal utility of  $W_t$ , is positive

no matter what  $\bar{W} - \Pi^t$  is, and this is true for all  $t$ . Moreover, since the decisionmaker is a risk averter (i)  $\Pi^t$  is semipositive for all  $t$  so that

$\bar{W} > \bar{W} - \Pi^t(\bar{W}, \tilde{Z})$  and (ii)  $U(W)$  is concave. But (i) and (ii) imply that

$[dU(W)]_{\bar{W} - \Pi^t} - [dU(W)]_{\bar{W}}$  is positive for all  $t$ .

Hence, we conclude that

$$(6.29) \quad d\pi_t^t > 0$$

and this is true for all  $t$ . But this exactly meets the condition for increasing risk aversion. In fact, it is stronger than the condition defining increasing risk aversion since the latter requires only that the strict inequality in (6.29) hold for one  $t$  so long as  $d\pi_t^t \geq 0$  for all  $t$ .

It is also important to note the generality of this proof. The preceding theorem and proof in no way restricted the class of risks  $\tilde{Z}$  that were considered, beyond the standard restriction to actuarially neutral risks. The multiperiod quadratic utility function was shown to be increasingly risk-averse for any vector  $\tilde{Z}$  of fair risks. This contrasts with the case of the proof of the sufficient conditions for multiperiod decreasing risk aversion that were presented in Theorem 2.2. In discussing sufficient conditions, it will be recalled, the only risks that the present author found he could consider were vectors of independent risks. The conditions in Theorem 2.2 were only proven to be sufficient for multiperiod decreasing risk aversion in the face of risks such that  $\tilde{Z} = \sum_{t=1}^T \tilde{Z}^t$  where each  $\tilde{Z}^t$  has zeros everywhere but in the  $t$ th element.

In conclusion, then, no matter whether one considers its use in a single-stage investment problem or in a multiperiod setting, the mean-variance approach seems unsuitable as the basis for a model of capital budgeting under risk. In each case its derivation rests upon assumptions about the decision-making unit's utility function and/or the probability distributions of the returns from investments. The latter assumptions almost completely force one into supposing a multivariate normal distribution for all cash inflows while the former entails assuming the decisionmaker's utility function is quadratic. Both of these suppositions have been shown to have implausible implications

either about the investment returns or about the decisionmaker's attitudes toward the utility of those returns and his behavior under risk.

#### 6.4.E. The Time-State-Preference Approach

The final theoretical suggestion in the literature that shall be considered is best introduced by considering an objection one of its advocates, J. Hirshleifer, has to the mean-variance approach. Define a "state of the world" as a complete world-environment for the decisionmaking unit. For example, two states of the world for an economy as a whole might be war versus peace; or health versus illness for an individual; or thriving business versus bankruptcy for a firm. Suppose the investor -- individual or corporate -- perceives the future as a set of three equally likely, mutually exclusive and exhaustive states of the world. One and only one of states A (a year of great losses for the firm), B (an average year for the firm), and C (a highly profitable year) will be realized. Two alternative sets of investments are contemplated, the vector of payoffs of the first being (3,2,1) and that of the second being (1,2,3), where the first element indicates the return if state A (year of great losses) occurs, the second element if B (average profits) occurs, and the third if C (a highly profitable year) is realized.

As one can easily calculate, the mean of the first portfolio's return equals that of the second, and the variances of the two are also equal. Using a mean-variance approach, the decisionmaker would then be indifferent between the two sets of investments. "But," writes Hirshleifer, "we have no right to assume that an investor would be indifferent between the two prospects. The nature of the world-environments A, B, and C might be such that he prefers the distribution biased toward wealth in state A over that biased toward wealth in state C. Here the ordering on A, B, and C of the elements of the distribution cannot be neglected; a distribution that is two-parameter disregarding ordering turns

out to be insufficiently specified, for preference ranking purposes, by the mean and standard deviation."<sup>50</sup> In the example used here, the firm would probably prefer the first portfolio of investments over the second since the former would return more when the rest of the corporation's operations were doing badly. The second portfolio would not greatly improve the situation in leaner years but would only serve to make highly profitable years more profitable. Mean and variance alone might well be an inadequate guide to portfolio selection in this situation.

The alternative approach advocated by Hirshleifer and pioneered by Arrow is the "time-state-preference" approach.<sup>51</sup> While the underlying objects of choice in the mean-variance approach are the expected return and variability of return of a portfolio, the objects of choice in the "time-state-preference" approach are postulated "to be contingent consumption opportunities, in alternative possible states of the world."<sup>52</sup> For instance, the firm in the example was to receive a return of 3 units from portfolio A if it had a year of great losses but only a 1-unit return from that set of investments if it had a highly profitable year in its other operations. An individual stockholder's contingent consumption opportunities in a given year from a firm he owns part of might, for example, be the regular dividend of that corporation if it has an average year, the regular dividend plus a bonus if the firm does very well, and nothing at all if the firm actually suffers losses in that year. As a last example, a farmer's consumption opportunities -- assuming fixed prices for the commodities he sells -- might depend on whether the year is one of adequate rainfall (in which case his consumption opportunities will be high) or one of drought (in which case he will have to restrict his consumption severely). With the introduction of these

---

<sup>50</sup>Hirshleifer (1965, p. 522).

<sup>51</sup>See Arrow (1964); Hirshleifer (1964, 1965, 1966).

<sup>52</sup>Hirshleifer (1965, p. 523).

conditional commodities as the fundamental objects of choice, the advocates of this approach are able to identify the problem of value under risk with the traditional problem of value under certainty. The two problems take exactly the same form when these contingent claims are introduced and, as Hirshleifer shows, the time-state-preference model can be viewed as a natural generalization of Fisher's choice-theoretic system.

An interesting idea to which the time-state-preference approach leads is the possibility that "the Neumann-Morgenstern function itself may not be independent of state -- since different states are not mere gamble outcomes like Black or Red at roulette, but rather may represent objective differences in one's external or internal context for choice."<sup>53</sup> The decisionmaker's utility function may not be unique. It may change depending on the state in which he finds himself. This variation might, for example, be attributed to the presence of "nonpecuniary income," positive or negative, in one or more of the possible states contemplated by the decisionmaker. Hirshleifer's objection to the mean-variance approach in the case of the two portfolios -- (3,2,1) and (1,2,3) -- rests on such asymmetrical preferences of the decisionmaker with respect to the different possible states.

The time-state-preference approach is a very interesting and important theoretical contribution. In the present context, where the basic utility function to be optimized is management's perception of the owners' utility, it is difficult to see how the influence of such different states could be given an important role. This is not to say that all alternatives can be described entirely in terms of (monetary) consumption income. The same consumption income may be valued differently depending on the circumstances under which it is received, the attendant nonpecuniary income, and so on. But it

---

<sup>53</sup>Hirshleifer (1964, p. 82).

seems legitimate in the present context of a corporation's capital-budgeting problem to abstract from such complications, especially since the present model takes as given the certain cash throw-offs from the firm's ongoing operations. It will be assumed that there is one basic utility function  $U(W_1, W_2, \dots, W_T)$  that is invariant with all states of the world. Behind this statement, of course, is the implicit assumption that the firm's position does not alter so radically during any one period as to make this uniqueness supposition completely untenable; for example, the corporation does not go bankrupt.

#### 6.5. A Closing Word

This completes our examination of the major previously suggested approaches -- both the more practical and the more theoretical -- to the problem of budgeting capital in an environment of risk. The shortcomings of the various procedures, excepting the time-state-preference approach from any general statement about defects in previous discussions of the question, have been examined in some detail. In the next chapter, we turn to a reconsideration, in light of the new risk environment, of the programming model presented earlier for budgeting capital in a certain world. The shortcomings of the previous proposals discussed in this chapter will help point the way to a better model of capital budgeting under risk.

## CHAPTER 7

### THE CAPITAL-BUDGETING PROBLEM UNDER RISK: AN AXIOMATIC APPROACH TO A PROGRAMMING MODEL

#### 7.1. Introduction

Having specified the risk environment within which the firm makes its capital-budgeting decision and having examined previous suggestions for making this decision, the present chapter considers an alternative procedure for solving the capital-budgeting problem in the risk and imperfect-capital-market environment that has been described. First, recall that in Section 6.2 it was argued that the correct approach to capital budgeting in this setting is a programming approach. Second, recall also that the assumptions about the nature of the risk environment confronting the firm, as described in that section, ensure that the constraint set the corporation faces is the same under risk as under certainty. The constraints of the programming model in (5.19) thus constitute the constraint set of the present programming model of capital budgeting under risk.

The objective function of such a model, is, however, a different matter. The risks in the environment being considered center upon the cash inflows from the projects, the period-by-period payoffs of the investment proposals. And these are the essential elements of the objective function of the capital-budgeting model developed in this study. The case presented earlier for an objective function that is a nonlinear utility function, namely, management's perception of the owners' utility with the amounts available for withdrawal



in the several periods serving as proxies for the true increases in stockholders' consumption possibilities, is as convincing under risk as under certainty. But now the arguments of that function, consisting of the payoffs from the newly undertaken projects and the withdrawals of funds generated by previously owned resources, are stochastic.

The question that must be answered is whether the presence of risk in this form imposes any further restrictions on the shape of the utility function that is at the center of our model of capital budgeting? And, of course, if risk does necessitate more detailed specification of the function, in addition to general nonlinearity, what are these additional restrictions? Proceeding in an axiomatic fashion, a set of desiderata will be specified that a utility function should possess if it is to serve as the basis for capital budgeting under risk. Then a utility function satisfying these conditions will be presented and the additional restrictions imposed on  $U(W_1, W_2, \dots, W_T)$  by the presence of risk will become clear. The implications of this utility function for attitudes toward the different characteristics -- expected return, variance of return, skewness of the distribution of return -- of the investment projects are discussed. Finally, the objective function presented and the constraint set are put together to yield a programming model of capital budgeting under risk.

## 7.2. The Expected-Utility Theorem as the Basis for Behavior Under Risk

As the fundamental basis for behavior under risk, I take the expected-utility theorem. The decisionmaker pursues the policy that maximizes the expected value of his numerical-valued utility function. The objective function for capital budgeting under risk is then

(7.1) Maximize

$$E\{U(W_1, W_2, \dots, W_T)\} = \int_A \dots \int U(W_1, W_2, \dots, W_T) f(W_1, W_2, \dots, W_T) dW_1 dW_2 \dots dW_T$$

where  $f(W_1, W_2, \dots, W_T)$  is the multivariate frequency function of the  $T$  stochastic amounts available for withdrawal and  $A$  is the region of integration defined by the ranges of the  $W_t$ 's.

A potential objection to this approach and a proper response are both given in the classic article by M. Friedman and L. J. Savage, although their hypothesis is actually a more specific version of (7.1), to which I shall take exception for other reasons. They write:

An objection to the hypothesis ... that is likely to be raised ... is that it conflicts with the way human beings actually behave and choose. Is it not patently unrealistic to suppose that individuals consult a wiggily utility curve ..., that they know the odds ..., that they can compute the expected utility ..., and that they can base their decision on the size of the expected utility?  
[to which they reply,]

While entirely natural and understandable, this objection is not strictly relevant. The hypothesis does not assert that individuals explicitly or consciously calculate and compare expected utilities. Indeed, it is not at all clear what such an assertion would mean or how it could be tested. The hypothesis asserts rather that, in making a particular class of decisions, individuals behave as if they calculated and compared expected utility and as if they knew the odds.<sup>1</sup>

Although no attempt is being made here to catalogue all the objections to the expected-utility theorem or to the Von Neumann-Morgenstern axioms to which it is equivalent, one other potential objection ought to be mentioned. It is the fact that the expected-utility theorem abstracts from phenomena as the "love of gambling," the "pleasure of suspense" and the like.<sup>2</sup> There is no

---

<sup>1</sup>Friedman and Savage (1948, p. 86).

<sup>2</sup>See Archibald (1959) for a statement of this objection. Archibald proposes a test of the expected-utility theorem that he believes the theorem would fail and that he believes demonstrates the need for introducing gambling preferences. J. Marschak (1950) presents a set of axioms, different from the set presented by Von Neumann and Morgenstern, that also leads to the maximization of expected utility by the "rational man." In the article, Marschak discusses the impossibility of comprehending "the pleasure of gambling" or "the pleasure of suspense" within the expected-utility-theorem framework.

difference, for example, in the utility of a marginal dollar of actual withdrawal when it is generated by a project that had a 99 per cent chance of paying that dollar rather than by a project that had only a 5 per cent probability of yielding that return. Similarly, differences among projects in terms of the suspense of undertaking them or curiosity about their result and so on have no effect on the utility value of the dollars they actually return.

With respect to this objection, I am in complete agreement with the position taken by Markowitz.

This writer believes that the arguments in favor of the expected utility maxim are quite convincing, especially for its application in areas such as portfolio selection [among which I would number capital budgeting]. The maxim has to be stretched, perhaps intolerably, to apply to the making of decisions in which surprise and the fun of gambling are important motivations. These, however, are not important objectives for the direction of a machine in the allocation of large amounts of other people's money.<sup>3</sup>

Markowitz's "machine" is his alternative to what is more commonly called the "rational man": the decisionmaker who, given his limited information and powers, makes no mistakes in arithmetic or logic in selecting an action to achieve his clearly defined objective. It is, indeed, such a "machine" for which the present study is trying to derive directions in allocating large amounts of corporate stockholders' money among alternative capital-investment projects. In such a case, I would argue, surprise and the fun of gambling are not important goals and should not be considered.

### 7.3. The Axioms

#### Axiom 1

Given this expected-utility basis for behavior under risk, the first property a utility function that is to serve as the basis for the capital-budgeting objective function ought to possess is positive marginal utility for

<sup>3</sup>Markowitz (1959, p. 210).

each period's consumption income. No matter what the levels of consumption incomes in the T periods, having a dollar more in any period is preferred to not having it. All first partial derivatives of the utility function must, consequently, be positive.

### Axiom 2

Second, the utility function must be bounded from above and from below. This condition, which is often overlooked, is required if the expected-utility theorem is to be used. Otherwise one falls into the trap of a version of the St. Petersburg paradox first discussed by D. Bernoulli in 1732.<sup>4</sup> The need for boundedness of the utility function if such difficulties are to be avoided was first pointed out by K. Menger who proved that "For any evaluation of additions to a fortune by an unbounded [utility] function, there exists a game related to the Petersburg Game in which the subjective expectation of the risk-taker on the basis of that value [utility] function is infinite."<sup>5</sup> In his recent papers on risk, Arrow has strongly emphasized this requirement imposed by the use of the expected-utility theorem.<sup>6</sup>

### Axiom 3

The third axiom states that the decisionmaking unit is a strict risk averter. It will prefer its status quo position to accepting a set of fair or actuarially neutral risks  $\tilde{Z}$ , that is, to accepting a set of gambles in which the expectation involves no change from the status quo,  $E(\tilde{Z}) = 0$ . This risk-aversion axiom is equivalent, as was proven in Chapter 2, to requiring that the utility function  $U(W_1, W_2, \dots, W_T)$  in (7.1) be strictly concave. Including this axiom

---

<sup>4</sup>Bernoulli (1738, p. 35). For a further discussion of the St. Petersburg paradox see Menger (1934).

<sup>5</sup>Menger (1934, p. 218).

<sup>6</sup>Arrow (1963, p. 26; 1965, pp. 25-27). Markowitz also recognized the importance of a utility function being bounded from above and from below when applying the expected-utility theorem. See Markowitz (1952, p. 154).

thus excludes the possibility that the utility function is linear in the consumption incomes of the several periods.

A linear utility function was rejected as a possibility even in the world of certainty. When risk is introduced it fails a fortiori as a candidate for the capital-budgeting objective function. If the utility function were linear,

$$(7.2) \quad U(W_1, W_2, \dots, W_T) = \sum_{t=1}^T u_t W_t,$$

the  $u_t$ 's being constants, the maximand under risk would be

$$(7.3) \quad E\{U(W_1, W_2, \dots, W_T)\} = \sum_{t=1}^T u_t E(W_t) \\ = \sum_{t=1}^T u_t \left[ \sum_{i=1}^n E(a_{ti}) y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n E(a_{tij}) x_{ij} + (1+r_L) R_t \right].$$

Maximizing the expected value of a linear utility function is thus equivalent to maximizing the expected value of the investment program's period-by-period returns, weighted by the constant  $u_t$ 's. But this ignores all aspects of the risk involved, for example, the dispersion of returns, as it weights extreme possible outcomes symmetrically, and hence allows neither for risk aversion nor for risk loving. Unless the decisionmaker is completely indifferent to the presence of risk, the linear utility function is unsuitable for use as the basis of his objective function. By postulating risk aversion for the capital-budgeting decisionmaker, the present study has excluded the linear utility function as a possibility.

Recent empirical work seems to support this rejection of the linear utility function. P. E. Green writes the following about his attempts to estimate utility functions for middle-management personnel in a large company.

However, it is of interest to note that even within this small-scale investigation, with its associated design difficulties, for this group of respondents at least, utility functions do not appear to be linear throughout the whole range of payoffs. If true, this finding raises the question that proximate measures of utility, e.g., cash flow, return on investment, etc., may be inadequate approximations to capital budgeting problems even in cases where the resources of the firm are not unduly strained by the alternatives being evaluated.<sup>7</sup>

Another more-or-less armchair empiricist writes of British firms that "The conservative [concave] utility function appears to be virtually universal for 'serious' decision-making situations, although other forms are possible."<sup>8</sup> And he goes on to define the "Capital Investment Problem" to be one which involves resources of such magnitude that the assumption of a linear utility function is unsatisfactory.

While the case may be relatively clear-cut for rejecting the assumption of linearity of the utility function, postulating risk aversion throughout the range of relevant consumption incomes has not been attractive to all theorists who have discussed decisionmaking under risk. In particular, Friedman and Savage, in the article referred to earlier, accept the expected-utility theorem and hypothesize a doubly-inflected single-stage utility function as the one in Figure 7.1.<sup>9</sup> The curve is first concave, then convex, and then becomes concave again. The individual is thus a risk averter for an initial range of consumption income (up to  $W'$ ), then he becomes a risk lover over a second range, and finally, after reaching a second critical level of income ( $W''$ ), he becomes risk-averse again. They infer this shape of the utility function from the following phenomena: "(a) low-income consumer units buy, or are willing to buy, insurance; (b) low-income consumer units buy, or are willing to buy, lottery

---

<sup>7</sup>Green (1963, p. 40).

<sup>8</sup>Adelson (1965, p. 32).

<sup>9</sup>Friedman and Savage (1948, p. 85, Figure 3).

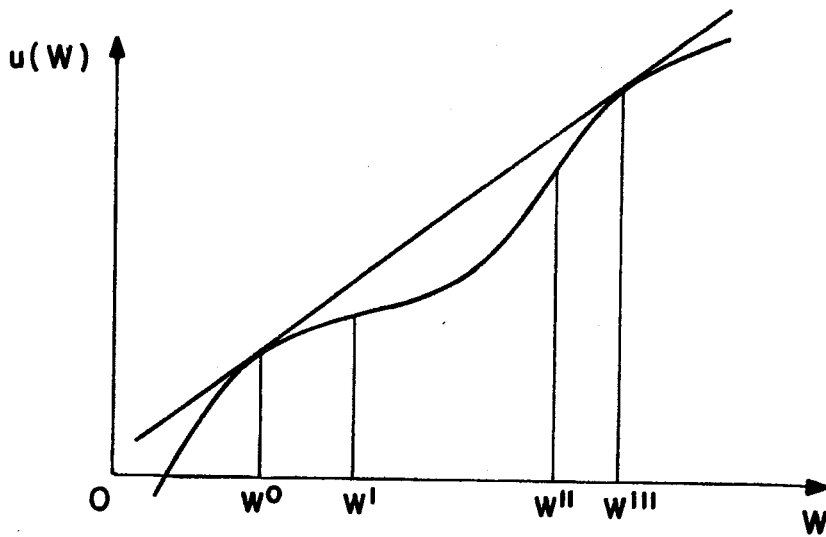


FIGURE 7.1  
The Friedman-Savage Utility Function

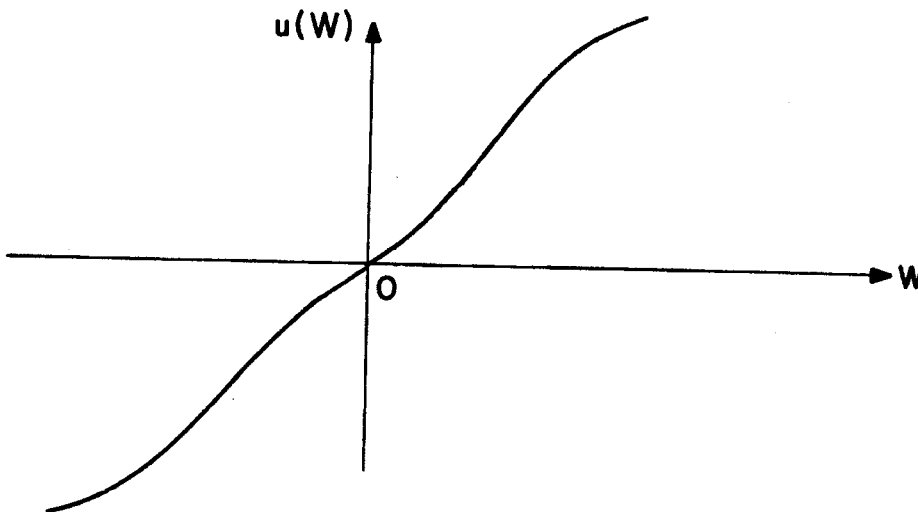


FIGURE 7.2  
The Markowitz Utility Function

tickets; (c) many consumer units buy, or are willing to buy, both insurance and lottery tickets; (d) lotteries typically have more than one prize."<sup>10</sup> Their suggested interpretation is that the concave segments correspond to socioeconomic classes while the risk-loving convex segment is a transitional stage between the lower and higher classes. The majority of the consumer units, they hypothesize, have incomes that place them in either the lower or the upper class.

Several years after the Friedman-Savage article appeared, Markowitz suggested an alternative shape for the utility function that he asserted explained what the Friedman-Savage hypothesis explained but avoided the contradictions of common observation that he found in their hypothesis. First, he finds strange the behavior implied for middle-class people because "We do not observe persons of middle income taking large symmetric bets," as the Friedman-Savage hypothesis would assert they do. In fact, "We expect people to be repelled by such bets. If such a bet were made, it would certainly be considered unusual and probably irrational."<sup>11</sup> Second, the behavior implied for an individual with wealth slightly below  $W'''$  in Figure 7.1, "an almost rich" person, contradicts common observation. Such a person would supposedly, according to the curve in Figure 7.1, like nothing more than to engage in a bet which if won would raise him to  $W'''$  while if lost would plummet him to  $W^0$ . "He would be willing to take a small chance of a large loss for a large chance of a small gain. He would not insure against a loss of wealth to  $W^0$ . On the contrary he would be anxious to underwrite insurance. He would even be willing to extend insurance at an expected loss to himself!"<sup>12</sup> Finally, the Friedman-Savage hypothesis implies that a person with wealth less than  $W^0$  or more than  $W'''$  would never take a fair (and

---

<sup>10</sup> Ibid., p. 95.

<sup>11</sup> Markowitz (1952, p. 152).

<sup>12</sup> Ibid., pp. 152-153. The notation for the wealth levels has been changed from Markowitz's labeling to conform to the notation of the present study.



hence never an unfair) bet. This, says Markowitz, is implausible in light of the fact that both poor and rich people do gamble.

In place of the Friedman-Savage hypothesis, Markowitz suggests the utility curve in Figure 7.2.<sup>13</sup> This function has three inflection points with the middle one defined to be at the "customary" level of wealth. This wealth level is equal to present wealth except in the case of windfall gains or losses. Note that this setting of the origin at customary wealth means the curve slides up and down as the individual's wealth level changes.

The decisionmaker is first risk-loving, then risk-averse, then risk-loving again, and finally risk-averse once more. He goes through two phases on each side of his present wealth level with the risk-loving (convex) segment preceding the risk-averse (concave) one in each instance. The inflection points get further apart (or at least get no closer) as the individual's wealth increases or, equivalently, as individuals of greater wealth are considered. Since people avoid symmetric bets -- at least Markowitz observes that they do -- the curve is such that  $|U(-W)| > U(W)$  where  $W > 0$ . That is, it falls more steeply to the left of present wealth than it rises to the right. Markowitz summarizes the implications of his hypothesized utility curve as follows:

Thus we see that the hypothesis is consistent with both insurance and lotteries, as was the F-S [Friedman-Savage] hypothesis. We also see that the hypothesis avoids the contradictions with common observations [namely, the existence of classes of people who prefer large symmetric bets to any other bets or people who desire to become one-man insurance companies, even at an expected loss] to which the F-S hypothesis was subject.<sup>14</sup>

The difficulties with the Friedman-Savage hypothesis, however, are too fundamental to be circumvented by proliferating hypothesized points of inflection.

<sup>13</sup> Ibid., pp. 154-155. Figure 7.2 appears as Figure 5 on p. 154 of Markowitz's article.

<sup>14</sup> Ibid., p. 155.

As Hirshleifer puts it, "As in the case of Ptolemy's geocentric system, when it becomes necessary to incorporate such ad hoc 'epicycles' to save the phenomena, it is time for a new conception."<sup>15</sup> While low-income families can be observed to place long-shot bets that might catapult them out of poverty or to purchase insurance or to do both, thereby verifying the behavior implied for them by the curve in Figure 7.1, the hypothesis fails to predict accurately the behavior of the middle-income and high-income classes. For example, the Friedman-Savage curve implies that wealthy individuals at the lower end of the high-income scale (that is, with wealth close to but above  $W''$ ) will avoid purchasing insurance or will accept bets that offer the strong possibility of a small gain and a weak possibility of a large loss. This is not commonly observed.

The real failure of the Friedman-Savage hypothesis resides, however, with its middle convex segment, as Markowitz noted. The middle class, according to the utility curve in Figure 7.1, would be risk plungers of an extreme sort. Hirshleifer summarizes the issue excellently.

They would stand ready, at any moment, to accept at fair odds a gamble of such a scale as to thrust them out of the convex segment and into (depending on the outcome) the poor-man or rich-man class. In addition, ... it is the individual at the upper end of the low-income segment who is most inclined to take long-shot bets, and the individual at the lower end of the high-income segment who is most inclined to take short-odd bets. Thus, the model would have us believe, the solid risk-avoiders of our society are only the poorer poor, and the richer rich. Aside from the notorious lack of direct confirmation of these assertions, it is of interest to note how they conflict with observed stability patterns of the various income classes. With behavior as postulated, the middle ranks of incomes would be rapidly depopulated, the end result being a U-shaped distribution piled up at the extremes. Needless to say, this is not observed.<sup>16</sup>

---

<sup>15</sup>Hirshleifer (1966, p. 261, footnote 9).

<sup>16</sup>Ibid., pp. 260-261.

An attempt to build into the Friedman-Savage model an explanation of the observed rarity of possibly impoverishing gambling in the middle-income and upper-income classes is unsatisfactory. One could assert that the convex segment in Figure 7.1. is very small. But then risk aversion predominates and the weight of the Friedman-Savage argument is lost. Alternatively, one could resort to hypothesizing, as Markowitz did, a small risk-loving range but one that moves with the individual's wealth so that he is always prepared to gamble a small amount. The latter approach is open, as Hirshleifer points out, to the charge of incorporating "epicycles" into the equivalent of the geocentric system.

In addition, if one puts faith in some recent experiments performed by M. Yaari, both the Friedman-Savage hypothesis and the Markowitz hypothesis must be rejected. Yaari tests the hypothesis that the set of all bets a decision-maker is willing to accept on a given event is convex. That is, "Consider a family of mutually exclusive and exhaustive events on some sample space. A gamble  $\langle x_1, \dots, x_n \rangle$  is a contract which promises its owner  $x_1$  dollars if  $E_1$  occurs,  $x_2$  dollars if  $E_2$  occurs, and so on."<sup>17</sup> The convexity hypothesis means that the set of all gambles  $\langle x'_1, \dots, x'_n \rangle$  preferred to or indifferent to any given gamble  $\langle x_1, \dots, x_n \rangle$ , preference or indifference being determined by the decision-maker, is convex.<sup>18</sup> From his admittedly crude experiments, Yaari concludes that "the hypothesis that the set of all accepted bets is convex is borne out to a large extent. Furthermore, the convexity of the set of accepted bets is entirely consistent with gambling, with insurance, or with a mixture of the two."<sup>19</sup> The maximization of expected utility when the utility function is of either the Friedman-Savage or Markowitz type, that is, contains "dents," implies, on

---

<sup>17</sup>Yaari (1965, p. 279).

<sup>18</sup>Recall that a set is convex if for any two points in it, the line segment joining them is also in the set.

<sup>19</sup>Yaari (1965, p. 278).

the other hand, that the set of accepted bets need not be convex. If one accepts Yaari's verification of the convexity postulate, one cannot accept either the Friedman-Savage or Markowitz hypotheses.

In postulating general risk aversion, the present study therefore seems to stand on firm ground. At the same time, the possible simultaneous holding of lottery tickets and insurance policies is not excluded. The assumption is that with respect to major financial decisions (what Hirshleifer calls "wealth-oriented" activities) the decisionmaking unit -- and the ownership whose utility function is involved -- is risk-averse. On the other hand, in more casual or entertaining endeavors (what Hirshleifer calls "pleasure-oriented" activities), the decisionmakers or the people whom they represent may engage in a modest amount of gambling. Risk aversion predominates with respect to wealth-oriented activities while a moderate amount of pleasure-oriented gambling (gambling as a consumption good) occurs all along the income scale. "All this is not to say that we never observe wealth-oriented gambling, but rather that it is not a sufficiently important phenomenon to dictate the main lines of our theory of risk."<sup>20</sup>

#### Axiom 4

The fourth condition set for a utility function for capital budgeting under risk is that of decreasing risk aversion. The decisionmaker is assumed to want to pay less (in a sense made specific in Chapter 2) for insurance against a given vector of risks as his asset vector increases. The justification for this postulate has been given at length in Section 2.2 and somewhat more briefly in the discussion of the mean-variance approach in Section 6.4.D. One should recall at this time that Theorem 6.1 proved that the multiperiod quadratic

---

<sup>20</sup>Hirshleifer (1966, p. 262). This rationalization of the simultaneous presence of risk aversion with respect to wealth-oriented decisions and gambling for pleasure was developed by the present author before reading Hirshleifer's excellent article. Reading the latter, however, helped to clarify some thoughts and helped to improve the present exposition.

utility function was everywhere increasingly risk-averse. Requiring decreasing risk aversion of the utility function that is to serve as the basis of the capital-budgeting objective function thus excludes the quadratic utility function as a possibility.

Axiom 5

The final requirement set upon the utility function  $U(W_1, W_2, \dots, W_T)$  is of a somewhat different nature from the four preceding ones. This last property borders on the realm of analytical convenience as opposed to behavior description or rationality for its justification. Nevertheless, insofar as the problem would be extremely difficult if not analytically impossible to solve without this restriction, it seems reasonable to believe that actual decision-makers could not make their decisions within a framework in which this property did not obtain. Perhaps this will be clearer after the restriction has been stated. It is that the utility function not impose upon us the task of having to specify fully the frequency function  $f(W_1, W_2, \dots, W_T)$  of the random variables  $W_t$ ,  $t=1, \dots, T$ .

The problem is that specifying this joint frequency function involves specifying the joint frequency function of the T sums

$$(7.4) \quad W_t = \sum_{i=1}^n a_{ti} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n a_{tij} x_{ij} + (1+r_L)R_t, \quad t=1, \dots, T.$$

But each of these sums is an unknown random variable. It itself depends on a subset of the stochastic parameters  $a_{ti}$  and  $a_{tij}$ , with the random variables to be included in the subset only being determined simultaneously with the optimal solution. The optimization process would be, to say the least, extremely difficult if the full unknown frequency function  $f(W_1, W_2, \dots, W_T)$  had to be included in the objective function. The difficulties involved are similar

to those encountered when one attempts to solve a programming problem with stochastic elements in the constraint matrix, taking account of contingency plans for the decisionmaker. The reader will recall the pessimistic pronouncement by Hadley as to the possibility of locating a global optimum in such a problem.

If, in contrast, the utility function only required, for example, that a small number of low-order moments of the distribution of the  $W_t$ 's be specified, the situation would be very different. The individual and joint moments of the  $W_t$ 's are easily expressed in terms of the (assumed) known individual and product moments of the component random variables -- the  $a_{ti}$ 's and  $a_{tij}$ 's -- and the unknown decision variables, the  $y_i$ 's and  $x_{ij}$ 's. The values actually taken on by the various moments of the  $W_t$ 's would be determined only when the optimal  $y_i$ - and  $x_{ij}$ -values were found. But being able to express the objective function in (7.1) in terms of the lower order moments of the stochastic payoffs and the decision variables alone would render the optimization problem solvable.

This fifth property is most easily met by restricting consideration to polynomial utility functions. Such functions permit the reduction of the objective function of the capital-budgeting problem under risk to a function of the decision variables and the moments of the probability distributions of the individual projects' returns and cash-inflow interactions. This is true because any term that is summed in a polynomial utility function in the  $W_t$ 's can be written as  $W_1^{\eta_1} W_2^{\eta_2} \dots W_T^{\eta_T}$  where the  $\eta_t$ 's are nonnegative integers. But, then, when the expectations operator is applied to the polynomial utility function -- that is, when the multiple integration in (7.1) is performed -- the typical term will be  $E(W_1^{\eta_1} W_2^{\eta_2} \dots W_T^{\eta_T})$ . This last expression is an individual moment about the origin if all but one of the  $\eta_t$ 's are zero, for example,  $E(W_2^3)$ . Or it is a joint or product moment about the origin if more than one  $\eta_t$  is nonzero,

for example,  $E(W_1 W_2^2 W_T)$ . Substituting the expression in (7.4) for each  $W_t$  into  $E(W_1^{\eta_1} W_2^{\eta_2} \dots W_T^{\eta_T})$  will yield an expression in the individual and joint moments of the  $a_{ti}$ 's and  $a_{tij}$ 's and in the  $y_i$ - and  $x_{ij}$ -variables.

This result is also advantageous in terms of the goal of bringing together the consumption-alternatives approach of neoclassical capital theory and the project approach of more recent capital-budgeting theory. One can move from a utility function in consumption alternatives, as one would find in the former, to a project-oriented objective function, as would be found in the latter.

#### 7.4. A Utility Function for Capital Budgeting Under Risk

Applying the rule of Occam's Razor, it is desirable to meet the requirements set for a multiperiod utility function with a polynomial of lowest possible degree. The lowest-order polynomial function that satisfies the stipulated conditions is a cubic defined over restricted ranges of consumption incomes. That is, each  $W_t$  lies within a particular range, perhaps differing for different periods, which will be indicated shortly. Limiting the values of consumption incomes considered serves to bound the function and hence causes it to satisfy the second axiom. The ranges of increments to consumption income, that is, the ranges of gross returns and cash removals, to which attention is restricted are dictated by the decisionmaker's views as to the ranges he believes relevant and over which his attitudes are described by the axiom set presented.

While the first requirement -- that of positive marginal utility -- can be met by a linear function, the risk-aversion (concavity) axiom requires at least a second-degree function. The quadratic utility function is not appropriate, though, since it fails to satisfy the axiom of decreasing risk aversion. A cubic with suitably restricted parameter signs and values can, however, meet all the requirements, as will now be proven.

The utility function that will be presented will be additive in the utilities of the different periods. That is, it will be of the form

$$(7.5) \quad U(W_1, W_2, \dots, W_T) = \sum_{t=1}^T u_t(W_t) \quad ,$$

where each  $u_t(W_t)$  is a single-period utility function depending only on the consumption income in period  $t$ ,  $W_t$ . In Chapter 4, however, it was argued that a utility function that excluded the possibility of interactions between the dollars in different periods was overly restrictive. A multiperiod utility function suitable for use as the capital-budgeting objective function, the argument went, should allow for the possibility that the utility of having  $W_t$  dollars available for consumption in period  $t$  may not be independent of the purchasing power available in period  $t'$ . It is easier, however, to proceed by first presenting a separable cubic utility function that possesses the properties stipulated above. Then it can be shown how a cubic multiperiod utility function can incorporate interactions between different periods' consumption incomes, for a very general set of risks but not for all risks. A multiperiod cubic is thus not overly restrictive in the sense of requiring additivity of the utilities of the different periods.

In the development that follows this demonstration, though, the separable cubic function will be employed. The reason is quite simple. When the expected-utility theorem is applied and the  $W_t$ 's are replaced by their equivalents expressed in terms of the individual projects, the presence of interactions between the consumption incomes of different periods provides no additional analytical insights. Instead the existence of such interdependence serves only to complicate the notation when the project-formulation of the objective function is being developed. Hence, the strategy will be first to prove that a separable multiperiod cubic utility function satisfies the axioms; then to



prove that a multiperiod cubic can comprehend interperiod interactions among consumption incomes; but then in the interests of expositional simplicity, to employ the separable function in pursuing the analysis of the capital-budgeting problem.

7.4.A. A Mathematical Statement of Sufficient Conditions for Satisfying the Axioms

The demonstration that a suitably restricted multiperiod cubic utility function satisfies the set of axioms proceeds by presenting a mathematical statement of sufficient conditions for the fulfillment of the postulates by a function. As this statement, in particular the part concerning decreasing risk aversion, relies on material presented in Chapter 2, it is best to review some of the important notation here. First,  $W^t$  is a vector each of whose elements is zero except the  $t$ th which equals the  $t$ th element of the vector  $W = (W_1, W_2, \dots, W_T)$ , so that  $W = \sum_{t=1}^T W^t$ . Next,  $\tilde{Z} = (\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_T)$  is a vector of actuarially neutral independent risks so that  $E(\tilde{z}_t) = 0$  for all  $t$  and the distribution of  $\tilde{z}_t$  is unaffected by the realization of any other risk  $\tilde{z}_s$  for  $s \neq t$ . The vector  $\tilde{Z}^t$  is defined as one having zeros in every element except the  $t$ th which equals the random variable  $\tilde{z}_t$ , the  $t$ th element of the risk vector  $\tilde{Z}$ . As a result of the independence assumption concerning risks in different periods, we also have

$$\tilde{Z} = \sum_{t=1}^T \tilde{Z}^t .$$

The random variable  $\tilde{h}_t$  is defined to be  $U(W^t + \tilde{Z}^t)$ . That is,  $\tilde{h}_t$  is the stochastic variable giving the stochastic utility level that results when the decisionmaker faces the risk  $\tilde{Z}^t$  and the initial consumption-income vector is  $W^t$ . Finally, recall that  $U^{-1}[p, \bar{W}, t]$  is defined as the  $T$ -element vector  $W$  such that (i)  $U(W) = p$  and (ii)  $W_i$  equals  $\bar{W}_i$  for all  $i$  except  $i=t$ .

The mathematical statement of a set of sufficient conditions for  $U(W)$  to satisfy the axioms can now be given. These conditions are:

(7.6) Positive marginal utility:  $\frac{\partial U}{\partial W_t} > 0$  for all  $t$  and any  $W$ ;

(7.7) Boundedness of  $U(W)$ ;

(7.8) Risk aversion:  $U(W)$  is concave in  $W_1, W_2, \dots, W_T$  ;

(7.9) Decreasing risk aversion

(i)  $U\left(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t]\right)$  is convex in the  $\tilde{h}_t$  variables;

(ii)  $dU\left(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t]\right)$  is convex in the  $\tilde{h}_t$  variables where the differential is with respect to nonnegative increments in the  $T$  initial consumption incomes;

(iii) Setting all  $W_i = 0$  except  $W_t$  in  $U(W)$ ,

$$\frac{\partial^3 U}{\partial W_t^3} \cdot \frac{\partial U}{\partial W_t} \geq \left(\frac{\partial^2 U}{\partial W_t^2}\right)^2$$

and this is true for all  $t$ ;

(7.10) Polynomial.

7.4.B. A Separable Cubic Utility Function that Satisfies the Axioms

The proof that a suitably restricted multiperiod cubic utility function satisfies the axioms then consists of the proof of the following theorem.

THEOREM 7.1. If

$$(7.11) \left\{ \begin{array}{l} U(W) = \sum_{t=1}^T [ W_t^3 - 2k_t W_t^2 + (k_t^2 + g_t^2) W_t ] , \\ k_t > 0 \text{ and } g_t^2 > \frac{1}{3} k_t^2 \text{ for each } t \text{ and } \frac{2}{3} k_t - \frac{1}{3} \sqrt{3g_t^2 - k_t^2} \leq W_t < \frac{2}{3} k_t \\ \text{for each } t, \end{array} \right.$$

then  $U(W)$  satisfies conditions (7.6)-(7.10).

Comment: The function in the theorem is additive in the different periods' utilities as

$$(7.12) \quad U(W) = \sum_{t=1}^T u_t(W_t) \quad \text{with } u_t(W_t) = W_t^3 - 2k_t W_t^2 + (k_t^2 + g_t^2) W_t$$

subject to the restrictions on  $W_t$ . Each of these single-period utility functions is a specific type of cubic with one real root, which has been arbitrarily set equal to zero, and two complex roots,  $k_t + g_t i$  and  $k_t - g_t i$ . Setting the real root of each period's cubic to zero is equivalent to setting the zero on the utility scale so that  $U(0, 0, \dots, 0) = 0$ . Since the zero point on the utility scale is one of the degrees of freedom allowed in Von Neumann-Morgenstern utility theory, the present argument suffers no loss in generality by setting  $U(0, 0, \dots, 0) = 0$ .

PROOF: The polynomial condition in (7.10) is obviously fulfilled since the proposed utility function is a cubic in the consumption incomes of the several periods. Second, the utility function clearly satisfies the condition of boundedness in (7.7). With each  $g_t$  and  $k_t$  pair finite constants and the range of  $W_t$  restricted by them, the range of values each  $u_t(W_t)$ , and hence  $U(W)$ , can assume is also bounded. Now consider, in turn, each of the other members of the set of sufficient conditions.

(a) Positive marginal utility: (7.6)

$$(7.13) \quad \frac{\partial U}{\partial W_t} = 3W_t^2 - 4k_t W_t + (k_t^2 + g_t^2) \quad \text{which can be rewritten as}$$

$$(7.14) \quad \frac{\partial U}{\partial W_t} = 3(W_t - \frac{2}{3} k_t)^2 + (g_t^2 - \frac{1}{3} k_t^2) \quad .$$

Since  $(W_t - \frac{2}{3} k_t)^2 > 0$  and  $g_t^2 - \frac{1}{3} k_t^2 > 0$  from the restrictions in (7.11), it follows that  $\frac{\partial U}{\partial W_t} > 0$  for all  $t$ . Note that the proposed function exhibits positive marginal utility not only for the ranges to which the  $W_t$ 's are restricted,

but for all values of  $W_t$ , since  $g_t^2 - \frac{1}{3} k_t^2 > 0$  and  $(W_t - \frac{2}{3} k_t)^2$  is always nonnegative.

(b) Risk aversion: (7.8)

From (7.12)  $U(W) = \sum_{t=1}^T u_t(W_t)$ . But sums of concave functions are concave.<sup>21</sup>

Thus, if each  $u_t(W_t)$  is concave in  $W_t$ ,  $U(W)$  is concave in  $W = (W_1, W_2, \dots, W_T)$ .

$$(7.15) \quad u_t''(W_t) = 6W_t - 4k_t \quad .$$

Hence,  $u_t(W_t)$  is concave for all  $W_t < \frac{2}{3} k_t$ . This includes the entire range of  $W_t$  for which the function in (7.11) is defined. Therefore,  $U(W)$  is concave over the ranges of relevant consumption incomes, as required by (7.8).

(c) Decreasing risk aversion: (7.9)

Consider  $\tilde{h}_t = U(W^t + \tilde{Z}^t)$ . With  $W^t$  given,  $\tilde{Z}^t$  maps into a unique distribution for  $\tilde{h}_t$ . In addition, since the utility function exhibits positive marginal utility for consumption income in each period and since  $\tilde{Z}^t$  contains only one nonzero element, for  $W^t$  given,  $\tilde{h}_t$  maps into a unique risk vector  $\tilde{Z}^t$ . Hence,

$$(7.16) \quad \tilde{h}_t = U(W^t + \tilde{Z}^t) \quad \text{and} \quad W^t + \tilde{Z}^t = U^{-1}[\tilde{h}_t, W^t, t] \quad .$$

Similarly, for a given  $\tilde{Z}^t$ , each  $W^t$  maps into a unique distribution for  $\tilde{h}_t$  and by the positive-marginal-utility property, each distribution of  $\tilde{h}_t$  maps into a unique value of  $W^t$ . Note, moreover, that if attention is restricted to actuarially neutral risks,  $E(\tilde{Z}) = 0$ , then there is a one-to-one relationship between the distributions of  $\tilde{h}_t$  and the  $(W^t, \tilde{Z}^t)$  pairs. Each  $(W^t, \tilde{Z}^t)$  pair implies one and only one distribution of utility  $\tilde{h}_t$  and each distribution of utility when consumption income and risk exist in period  $t$  only implies one and only one  $(W^t, \tilde{Z}^t)$  pair.

As a last preliminary, suppose all  $W_i$ 's and all  $\tilde{z}_i$ 's are set equal to zero except for  $i = t$ . The decisionmaker with the utility function in (7.11)

<sup>21</sup>See, for example, Hadley (1964, pp. 85, 102).

will then have an horizon utility given by the tth component of that utility function. That is

$$(7.17) \quad \tilde{h}_t = U(W_t^t + \tilde{z}_t^t) = U(0, 0, \dots, W_t^t + \tilde{z}_t^t, 0, \dots, 0) = u_t(W_t^t + \tilde{z}_t^t) .$$

The previous remarks can then be summarized by saying that if

$E(\tilde{z}_t) = 0$ , then  $\tilde{h}_t = u_t(W_t^t + \tilde{z}_t^t)$  is a function with a unique inverse, that is,  $u_t^{-1}(\tilde{h}_t) = W_t^t + \tilde{z}_t^t$  is a function, not a correspondence.

$$(i) \quad U\left(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W_t^t, t]\right) = U\left(\sum_{t=1}^T [W_t^t + \tilde{z}_t^t]\right) \text{ from (7.16).}$$

From the definitions of  $W^t$  and  $\tilde{z}^t$  and from (7.12) it follows that this equals

$$U(W + \tilde{z}) = U(W_1 + \tilde{z}_1, W_2 + \tilde{z}_2, \dots, W_T + \tilde{z}_T) = \sum_{t=1}^T u_t(W_t + \tilde{z}_t) .$$

Therefore, from (7.17),

$$(7.18) \quad U\left(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W_t^t, t]\right) = \sum_{t=1}^T \tilde{h}_t .$$

Hence,  $U\left(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W_t^t, t]\right)$  is a linear function of the  $\tilde{h}_t$  variables; specifically, it is the sum of them. Therefore it is a convex function of the  $\tilde{h}_t$  variables since a linear function is convex.

$$(ii) \quad dU\left(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W_t^t, t]\right) = dU(W_1 + \tilde{z}_1, W_2 + \tilde{z}_2, \dots, W_T + \tilde{z}_T) ,$$

where the differential is with respect to nonnegative changes in the initial consumption incomes and we have used the same intermediate steps as in the proof of part (i) above. Therefore,

$$(7.19) \quad dU\left(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W_t^t, t]\right) = \sum_{t=1}^T \left(\frac{\partial U}{\partial W_t}\right)_{W + \tilde{z}} dW_t = \sum_{t=1}^T \left(\frac{du_t}{dW_t}\right)_{W_t + \tilde{z}_t} dW_t .$$

But from (7.13), this is equivalent to

$$(7.20) \quad dU\left(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t]\right) = \sum_{t=1}^T [3(W_t + \tilde{z}_t)^2 - 4k_t(W_t + \tilde{z}_t) + k_t^2 + g_t^2] dW_t .$$

From equation (7.17) and the statement following it,

$$(7.21) \quad W_t + \tilde{z}_t = u_t^{-1}(\tilde{h}_t) \quad \text{is a unique inverse function.}$$

Therefore, the equation in (7.20) may be rewritten as

$$(7.22) \quad dU\left(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t]\right) = \sum_{t=1}^T \{3[u_t^{-1}(\tilde{h}_t)]^2 - 4k_t u_t^{-1}(\tilde{h}_t) + k_t^2 + g_t^2\} dW_t .$$

A nonnegative linear combination of convex functions is, however, convex. Hence, since each  $dW_t \geq 0$  by assumption,  $dU\left(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t]\right)$  will be convex in the  $\tilde{h}_t$

variables if each term in braces is convex in the  $\tilde{h}_t$  of which it is a function.

The total differential of  $U$  with respect to nonnegative changes in the  $W_t$ 's will be a convex function of the  $\tilde{h}_t$  variables if each  $\frac{du_t}{dW_t}$  is convex in its respective

$\tilde{h}_t$ , that is, if  $\frac{d^2}{d\tilde{h}_t^2} \left( \frac{du_t}{dW_t} \right) \geq 0$ .

$$(7.23) \quad \frac{d^2}{d\tilde{h}_t^2} \left( \frac{du_t}{dW_t} \right) = \left( \frac{d^3 u_t}{dW_t^3} \right) \left( \frac{dW_t}{d\tilde{h}_t} \right)^2 + \left( \frac{d^2 u_t}{dW_t^2} \right) \left( \frac{d^2 W_t}{d\tilde{h}_t^2} \right) .$$

Since the inverse function in (7.21) is unique, for any arbitrary but given  $\tilde{z}_t$ ,

one obtains

$$(7.24) \quad \frac{dW_t}{d\tilde{h}_t} = \frac{1}{\frac{d\tilde{h}_t}{dW_t}} = \frac{1}{\frac{du_t}{dW_t}} \quad \text{and} \quad \frac{d^2 W_t}{d\tilde{h}_t^2} = - \frac{\frac{d^2 u_t}{dW_t^2}}{\left( \frac{du_t}{dW_t} \right)^3} .$$

Substituting these into (7.23) one finds

$$(7.25) \quad \frac{d^2}{dh_t^2} \left( \frac{du_t}{dW_t} \right) = \left[ \frac{1}{u_t'(W_t)} \right]^3 \{ u_t'''(W_t) u_t'(W_t) - [u_t''(W_t)]^2 \},$$

writing  $\frac{du_t}{dW_t}$  as  $u_t'(W_t)$ ,  $\frac{d^2 u_t}{dW_t^2}$  as  $u_t''(W_t)$ , and so on. Since  $u_t'(W_t) > 0$ ,

as shown in (7.14), it follows that one will have  $\frac{d^2}{dh_t^2} \left( \frac{du_t}{dW_t} \right) \geq 0$  if it is true that

$$(7.26) \quad u_t'''(W_t) u_t'(W_t) - [u_t''(W_t)]^2 \geq 0.$$

The requirement in (7.26) will be recognized as the condition for single-period decreasing risk aversion in the tth period. In addition, since the utility function that results if all  $W_i$ 's but  $W_t$  are set equal to zero in the utility function  $U(W)$  defined in (7.11) is  $u_t(W_t)$ , the condition in (7.26) is, in the present case, also the third part of the sufficiency condition in (7.9).

But (7.26) is satisfied by the proposed utility function over the ranges of consumption incomes for which it is defined. Substituting the appropriate expressions into (7.26), one obtains

$$(7.27) \quad 6(3W_t^2 - 4k_t W_t + k_t^2 + g_t^2) - (6W_t - 4k_t)^2 \geq 0.$$

This is equivalent to

$$(7.28) \quad \left( W_t - \frac{2}{3} k_t \right)^2 - \frac{1}{3} \left( g_t^2 - \frac{1}{3} k_t^2 \right) \leq 0.$$

The first restriction in (7.11) ensures that  $g_t^2 - \frac{1}{3} k_t^2 > 0$  so that  $-\frac{1}{3} \left( g_t^2 - \frac{1}{3} k_t^2 \right) < 0$ . This does not, however, suffice to guarantee that the sign of the entire expression in (7.28) is negative since  $\left( W_t - \frac{2}{3} k_t \right)^2 \geq 0$ .

But the inequality in (7.28) will be true if  $\frac{1}{3} \left( g_t^2 - \frac{1}{3} k_t^2 \right) \geq \left( W_t - \frac{2}{3} k_t \right)^2$  or

$$(7.29) \quad \frac{2}{3} k_t - \frac{1}{3} \sqrt{3g_t^2 - k_t^2} \leq W_t \leq \frac{2}{3} k_t + \frac{1}{3} \sqrt{3g_t^2 - k_t^2}.$$

Both of these inequalities are guaranteed by the restriction on the ranges of relevant consumption incomes in (7.11). In fact, the upper bounds of those ranges  $\frac{2}{3} k_t$ , are more restrictive than the right-hand inequality in (7.29) requires.

Since condition (7.26) obtains for the proposed utility function, each  $\frac{\partial U}{\partial W_t}$  is convex in its respective  $\tilde{h}_t$  variable. As a result, the total differential  $dU(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t])$ , being a nonnegative linear combination of functions convex in the  $\tilde{h}_t$  variables, is itself convex in those variables.

(iii) Setting all  $W_i = 0$  except  $W_t$  (for any  $t$ ) in the utility function defined in (7.11), one finds -- as a result of its separability expressed in (7.12) -- that the condition  $\frac{\partial^3 U}{\partial W_t^3} \frac{\partial U}{\partial W_t} \geq (\frac{\partial^2 U}{\partial W_t^2})^2$  is equivalent to condition (7.26). This condition has, however, just been proved true for the proposed utility function. Hence, condition (iii) of (7.9) is met.

In conclusion, it has been shown that the utility function given in (7.11), defined over the ranges of consumption incomes given there, and subject to the restrictions on its parameters indicated there, satisfies the set of sufficient conditions (7.6)-(7.10). It therefore satisfies the set of axioms for a utility function that is to serve as the basis for the capital-budgeting problem's objective function.

#### 7.4.C. A Nonseparable Cubic Utility Function that Satisfies the Axioms

Having proven that the additive multiperiod cubic in (7.11) meets the conditions stipulated, a utility function appropriate for use in making the capital-budgeting decision is at hand. Before going on to investigate the implications of that utility function for the decisionmaker's behavior and to derive the project formulation of the resulting objective function, let us consider an important digression. It will now be demonstrated that a multiperiod cubic utility function can incorporate certain interactions between



consumption incomes in different periods, at least over limited ranges of consumption incomes, and still satisfy the axioms. In particular, we prove the following theorem.

THEOREM 7.2: If

$$(7.30) \left\{ \begin{array}{l} U(W) = \sum_{t=1}^T [W_t^3 - 2k_t W_t^2 + (k_t^2 + g_t^2) W_t] + \sum_{t=1}^{T-1} m_t W_t W_{t+1} \\ \text{with } k_t, m_t \text{ positive for all } t ; \frac{1}{3} k_t^2 < g_t^2 < \frac{7}{9} k_t^2 \text{ for each } t; \\ W_t > 0 \text{ for each } t; \\ \frac{2}{3} k_t - \frac{1}{3} \sqrt{3g_t^2 - k_t^2} \cong W_t < \frac{2}{3} k_t \text{ for each } t, \text{ and} \\ \frac{4}{9} k_t - \frac{1}{9} \sqrt{7k_t^2 - 9g_t^2} < W_t < \frac{4}{9} k_t - \frac{1}{9} \sqrt{7k_t^2 - 9g_t^2} . \text{ for each } t, \end{array} \right.$$

then if each  $m_t$  is sufficiently small, the utility function  $U(W)$  satisfies conditions (7.6)-(7.10) for all "nonbankruptcy" risk situations, that is, for all wealth- and risk-vector combinations such that  $W_t + \tilde{z}_t$  is always positive for all  $t$ .

Comment: The function in the theorem is clearly not additive in the individual periods' utilities. It is, instead, the cubic utility function of Theorem 7.1, with a further restriction on each  $W_t$ , supplemented with single-period interactions between consumption incomes in successive periods. The purchasing power possessed today interacts positively with the consumption income available yesterday and that to be possessed tomorrow in determining the decisionmaker's horizon utility. It seems likely that if interactions between consumption incomes do exist, the strongest interdependence of a particular period's consumption income will be with that of the period immediately preceding it and with that of the one

immediately following it. The interaction that exists between successive periods' dollars might, however, be negative in some cases. Alternatively, the interdependence among different periods' consumption incomes might be much more complicated than the simple relationship indicated in (7.30). For example, in "real-world" utility functions one might find that the value of each period's consumption income depends on some weighted average of the purchasing power in the preceding  $T^*$  periods, or one might find a Duesenberry "ratchet effect."<sup>22</sup> All that is asserted in Theorem 7.2 is that there exists at least one set of reasonable interperiod dependences between consumption incomes that a multiperiod cubic utility function can comprehend while still satisfying the axioms set forth earlier.

Again, the zero point of the utility scale has been set at the origin of consumption-income space. The meaning of the characterization "sufficiently small" will become clear shortly and it will be seen that the question of how small "sufficiently small" is depends on the particular utility-function parameters and consumption-income ranges considered. In addition, the reasons for the further restrictions on the ranges of consumption incomes presented in (7.30) that did not appear in (7.11) will also become clear presently.

PROOF: The polynomial condition in (7.10) is again obviously satisfied. The fact that the range of each  $W_t$  is once again restricted by a set of finite constants guarantees that the range of values  $U(W)$  can assume is bounded so that (7.7) is satisfied. Now, consider each of the other members of the set of sufficient conditions.

(a) Positive marginal utility: (7.6)

---

<sup>22</sup>Duesenberry (1949).

$$(7.31) \quad \left\{ \begin{array}{l} \frac{\partial U}{\partial W_1} = 3W_1^2 - 4k_1W_1 + k_1^2 + g_1^2 + m_1W_2 \quad \text{for } t = 1 \\ \frac{\partial U}{\partial W_t} = 3W_t^2 - 4k_tW_t + k_t^2 + g_t^2 + m_{t-1}W_{t-1} + m_tW_{t+1} \quad \text{for } t=2, \dots, T-1 \\ \frac{\partial U}{\partial W_T} = 3W_T^2 - 4k_TW_T + k_T^2 + g_T^2 + m_{T-1}W_{T-1} \quad \text{for } t=T. \end{array} \right.$$

From the discussion of the additive cubic function, particularly equations (7.13) and (7.14), it is known that  $3W_t^2 - 4k_tW_t + k_t^2 + g_t^2 > 0$  for all  $t$ , as a result of the restrictions in (7.30). Hence, the marginal utilities of the different periods' consumption incomes in (7.31) will be positive so long as  $m_tW_t > 0$  and  $m_tW_{t+1} > 0$  for all  $t$ . With each  $m_t > 0$ , the utility function in (7.30) will thus certainly have positive marginal utility in each period if the only relevant consumption incomes are positive ones, that is, if  $W_t > 0$  for all  $t$ .<sup>23</sup> Since this is one of the restrictions on  $W_t$  given in (7.30), the utility function defined in (7.30) over the ranges of the  $W_t$ 's stated there satisfies condition (7.6).

(b) Risk aversion: (7.8)

A set of necessary and sufficient conditions for concavity of a function of many variables is that the naturally ordered principal minors of the Hessian of the function alternate in sign beginning with minus for the first-order principal minor. That is, the  $i$ th naturally ordered principal minor of the determinant whose elements are the second partial derivatives of the function must have the sign of  $(-1)^i$ . The Hessian of the utility function in (7.30) is

<sup>23</sup>In the case of the capital-budgeting problem discussed here, this is not a severe restriction for by the definition of  $W_t$  as the sum of the  $t$ th-period gross returns from new projects, the  $t$ th-period cash-inflow interactions from new projects, and the amount put aside for removal of funds in period  $t$ ,  $W_t$  will be positive unless there are strong negative cash interactions among projects.

$$(7.32) \quad H = \begin{vmatrix} 6W_1 - 4k_1 & m_1 & 0 & \dots & \dots & 0 \\ m_1 & 6W_2 - 4k_2 & m_2 & \dots & \dots & 0 \\ 0 & m_2 & 6W_3 - 4k_3 & m_3 & \dots & 0 \\ \vdots & 0 & m_3 & \dots & \dots & 6W_{T-1} - 4k_{T-1} & m_{T-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \vdots & \dots & \dots & m_{T-1} & 6W_T - 4k_T \end{vmatrix}.$$

Denoting the  $ij$ th element of  $H$  by  $H_{ij}$ , the conditions given in (7.30) imply that  $H_{tt} = 6W_t - 4k_t < 0$  by the restriction on  $W_t$ ;  $H_{t,t-1} = m_{t-1} > 0$  and  $H_{t,t+1} = m_t > 0$  by the restrictions on  $m_t$ ; and  $H_{ts} = 0$  for all  $s \neq t-1$  or  $t$  or  $t+1$ .

Every element of the determinant is zero except the elements on the main diagonal, each of which is negative, and the elements on the immediately sub- and super-diagonals, each of the elements on these latter diagonals being positive. If the values of the  $m_t$ 's are small enough, each naturally ordered principal minor can be reduced by elementary row operations to a triangular determinant with negative diagonal elements. That is, with "sufficiently small"  $m_t$ -values, the Hessian in (7.32) can be transformed into a triangular determinant the  $t$ th diagonal element of which has the same sign as  $6W_t - 4k_t$ , namely, negative. But the value of a triangular determinant is equal to the product of the elements on the main diagonal. With each main diagonal element negative, the sign of the  $i$ th naturally ordered principal minor would be that of  $(-1)^i$ , and the utility function in (7.30) would then be concave.

Exactly how small each  $m_t$  must be depends on how close the corresponding  $W_t$  actually comes to its upper bound  $\frac{2}{3} k_t$ . The closer the actual maximum value of  $W_t$  is to  $\frac{2}{3} k_t$ , the smaller the absolute value that  $6W_t - 4k_t$  can assume and the smaller must  $m_t$  be. The reason is not difficult to see. The conditions on the signs of the principal minors of  $H$  must hold for all possible  $W_t$ -values if the function is to be concave. The problem that may arise in reducing  $H$  to a triangular determinant is that some interaction(s) may be large enough relative to

the corresponding main-diagonal element(s) to make the sign of the diagonal element in the new triangular determinant positive. This would lead to violation of the conditions for concavity.

Consider, for example, the second naturally ordered principal minor in (7.32), namely,

$$\begin{vmatrix} 6W_1 - 4k_1 & m_1 \\ m_1 & 6W_2 - 4k_2 \end{vmatrix} .$$

To triangularize it, one wants to subtract  $\frac{m_1}{6W_1 - 4k_1}$  times the first row from the second which yields

$$\begin{vmatrix} 6W_1 - 4k_1 & m_1 \\ 0 & (6W_2 - 4k_2) - \frac{m_1^2}{6W_1 - 4k_1} \end{vmatrix} .$$

Recall that  $6W_1 - 4k_1 < 0$  and  $6W_2 - 4k_2 < 0$ . Therefore, if  $m_1$  is large enough it can make  $(6W_2 - 4k_2) - \frac{m_1^2}{6W_1 - 4k_1}$  positive. This would make the second naturally ordered principal minor negative, thereby violating the conditions for concavity. Clearly, the problem is most likely to arise when  $6W_2 - 4k_2$  is smallest in absolute value and when  $6W_1 - 4k_1$  is smallest in absolute value. The latter makes the situation worse because it increases the positive number being added, the former because it decreases the negative number to which we add. Hence, the smaller the minimum absolute values the  $6W_t - 4k_t$  terms can assume, the smaller must the  $m_t$ -parameters be.<sup>24</sup> And, if the  $m_t$ -values are sufficiently small, the naturally ordered principal minors of the Hessian will alternate in sign, beginning with minus, and  $U(W)$  in (7.30) will be concave in  $W$ .

---

<sup>24</sup>The curious reader may want to convince himself that the following set of interaction constants will yield a concave utility function of the form in (7.30). (continued)

(c) Decreasing risk aversion: (7.9)

The discussion preceding the proof of decreasing risk aversion for the separable utility function in (7.11) applies in full here. The functions

$$(7.16) \quad \tilde{h}_t = U(W^t + \tilde{Z}^t) \quad \text{and} \quad W^t + \tilde{Z}^t = U^{-1}[\tilde{h}_t, W^t, t]$$

define one-to-one mappings between the distributions of  $\tilde{Z}^t$  and  $\tilde{h}^t$  when  $W^t$  is given. Similarly, for a given  $\tilde{Z}^t$ , each  $W^t$  maps into a unique distribution for  $\tilde{h}_t$  and each  $\tilde{h}_t$  maps into a unique value of  $W^t$ . When attention is restricted, as it is here, to actuarially neutral risks so that  $E(\tilde{Z}) = 0$ , there is a one-to-one relationship between the distributions of  $\tilde{h}_t$  and  $(W^t, \tilde{Z}^t)$  pairs. Finally, it is easy to see that when all risks and consumption incomes are zero except the tth, the multiperiod utility of the decisionmaker with the utility function in (7.30) is given by the tth component of the additive cubic discussed earlier. Hence, if  $U(W)$  is the multiperiod utility function in (7.30),

$$(7.17) \quad \tilde{h}_t = U(W^t + \tilde{Z}^t) = u_t(W_t + \tilde{z}_t),$$

where  $u_t(W_t)$  is as defined in (7.12). If  $E(\tilde{z}_t) = 0$ , then,  $\tilde{h}_t = u_t(W_t + \tilde{z}_t)$  is a function with a unique inverse, that is,  $u_t^{-1}(\tilde{h}_t) = W_t + \tilde{z}_t$  is a function.

---

24 (continued)

$$m_t = \sqrt{\frac{(6W_t - 4k_t)^* (6W_{t+1} - 4k_{t+1})^*}{M}} \quad \text{for } t = 1$$

$$m_t = \sqrt{\frac{(6W_t - 4k_t)^* (6W_{t+1} - 4k_{t+1})^*}{M-t+2}} \quad \text{for } t \geq 2$$

where  $(6W_t - 4k_t)^* = \max(6W_t - 4k_t)$  and  $M$  is any integer greater than or equal to the number of periods,  $T$ . That is, with these interaction parameters, the Hessian in (7.32) based on the utility function in (7.30) can be reduced to a triangular determinant the main diagonal elements of which are all negative.

$$(i) \quad U\left(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t]\right) = U\left(\sum_{t=1}^T [W^t + \tilde{Z}^t]\right) = U(W + \tilde{Z})$$

from equation (7.16) and the definitions of  $W^t$  and  $\tilde{Z}^t$ . Using the statement of the utility function in (7.30) and the definition of  $u_t(W_t)$  in (7.12), one then obtains the equation in (7.33).

$$(7.33) \quad U\left(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t]\right) = \sum_{t=1}^T u_t(W_t + \tilde{z}_t) + \sum_{t=1}^{T-1} m_t(W_t + \tilde{z}_t)(W_{t+1} + \tilde{z}_{t+1}) .$$

Therefore, from (7.17) and the sentence following it,

$$(7.34) \quad U\left(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t]\right) = \sum_{t=1}^T \tilde{h}_t + \sum_{t=1}^{T-1} m_t u_t^{-1}(\tilde{h}_t) u_{t+1}^{-1}(\tilde{h}_{t+1}) .$$

The sum of convex functions is convex. But the first sum on the right-hand side,  $\sum_{t=1}^T \tilde{h}_t$ , is linear and hence convex in the  $\tilde{h}_t$  variables. Thus,  $U\left(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t]\right)$  will be convex in the  $\tilde{h}_t$  variables if the second sum on the right-hand side is convex. This will, in turn, be true if each component, each  $m_t u_t^{-1}(\tilde{h}_t) u_{t+1}^{-1}(\tilde{h}_{t+1})$  term, is convex. With each  $m_t > 0$ , this reasoning leads to the conclusion that  $U\left(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t]\right)$  will be convex in the  $\tilde{h}_t$  variables if

$$(7.35) \quad \theta_t = u_t^{-1}(\tilde{h}_t) u_{t+1}^{-1}(\tilde{h}_{t+1}) \quad \text{is convex in } \tilde{h}_t, \tilde{h}_{t+1} \text{ for each } t=1, \dots, T-1 .$$

The function  $\theta_t$  will, however, be convex if and only if

$$(7.36) \quad \frac{\partial^2 \theta_t}{\partial \tilde{h}_t^2} > 0 \quad \text{and} \quad \begin{vmatrix} \frac{\partial^2 \theta_t}{\partial \tilde{h}_t^2} & \frac{\partial^2 \theta_t}{\partial \tilde{h}_t \partial \tilde{h}_{t+1}} \\ \frac{\partial^2 \theta_t}{\partial \tilde{h}_t \partial \tilde{h}_{t+1}} & \frac{\partial^2 \theta_t}{\partial \tilde{h}_{t+1}^2} \end{vmatrix} > 0 .$$

Noting that  $u_t^{-1}(\tilde{h}_t)$  and its derivatives depend only on  $\tilde{h}_t$  and similarly  $u_{t+1}^{-1}(\tilde{h}_{t+1})$  and its derivatives depend only on  $\tilde{h}_{t+1}$ , one evaluates the partial derivatives in (7.36) and finds that the conditions there are equivalent to

$$(7.37) \quad \left\{ \begin{array}{l} \text{(i)} \quad u_{t+1}^{-1}(\tilde{h}_{t+1}) \cdot \frac{d^2}{d\tilde{h}_t^2} [u_t^{-1}(\tilde{h}_t)] > 0 \\ \text{(ii)} \quad \{u_t^{-1}(\tilde{h}_t) \frac{d^2}{d\tilde{h}_t^2} [u_t^{-1}(\tilde{h}_t)]\} \{u_{t+1}^{-1}(\tilde{h}_{t+1}) \frac{d^2}{d\tilde{h}_{t+1}^2} [u_{t+1}^{-1}(\tilde{h}_{t+1})]\} \\ \quad - \left\{ \frac{d}{d\tilde{h}_t} [u_t^{-1}(\tilde{h}_t)] \cdot \frac{d}{d\tilde{h}_{t+1}} [u_{t+1}^{-1}(\tilde{h}_{t+1})] \right\}^2 > 0 \end{array} \right. .$$

In order to demonstrate that the utility function in (7.30), subject to the restrictions given there, satisfies these two conditions, it is necessary to note the following relationships. Since  $u_t^{-1}(\tilde{h}_t)$  is a function, that is, since it is single-valued,

$$(7.38) \quad \frac{du_t^{-1}(\tilde{h}_t)}{d\tilde{h}_t} = \frac{1}{\frac{d\tilde{h}_t}{du_t(u_t^{-1}(\tilde{h}_t))}} = \frac{1}{\frac{d[u_t^{-1}(\tilde{h}_t)]}{du_t(u_t^{-1}(\tilde{h}_t))}} = \frac{1}{u_t'(W_t)}, \text{ and}$$

$$(7.39) \quad \frac{d^2 u_t^{-1}(\tilde{h}_t)}{d\tilde{h}_t^2} = - \frac{\frac{d^2 \tilde{h}_t}{d[u_t^{-1}(\tilde{h}_t)]^2}}{\left[ \frac{d\tilde{h}_t}{du_t^{-1}(\tilde{h}_t)} \right]^3} = - \frac{\frac{d^2 u_t(u_t^{-1}(\tilde{h}_t))}{d[u_t^{-1}(\tilde{h}_t)]^2}}{\left[ \frac{du_t(u_t^{-1}(\tilde{h}_t))}{du_t^{-1}(\tilde{h}_t)} \right]^3} = - \frac{u_t''(W_t)}{[u_t'(W_t)]^3},$$

and, finally,



$$(7.40) \quad u_t^{-1}(\tilde{h}_t) \frac{d^2 u_t^{-1}(\tilde{h}_t)}{d\tilde{h}_t^2} = u_t^{-1}(\tilde{h}_t) \frac{\frac{d^2 u_t^{-1}(\tilde{h}_t)}{d\tilde{h}_t^2}}{\left[ \frac{du_t^{-1}(\tilde{h}_t)}{d\tilde{h}_t} \right]^3} = -W_t \frac{u_t''(W_t)}{[u_t'(W_t)]^3} .$$

Substituting the appropriate expressions, as obtained from (7.17) and (7.39), into expression (7.37)(i), one obtains

$$(7.41) \quad (W_{t+1} + \tilde{z}_{t+1}) \left( - \frac{u_t''(W_t)}{[u_t'(W_t)]^3} \right) .$$

Since  $u_t''(W_t) = 6W_t - 4k_t < 0$  by the restriction on the range of  $W_t$ , and since  $u_t'(W_t) > 0$  as shown in the discussion of (7.14), the expression in (7.41) will be positive if  $W_{t+1} + \tilde{z}_{t+1} > 0$ . But this was one of the restrictions incorporated in the statement of Theorem 7.2, namely, that the utility function in question met the axioms for "nonbankruptcy" risk situations. By the latter is meant that no matter how the risk turns out the sum available for incrementing consumption income, namely,  $W_t + \tilde{z}_t$ , is positive in every period. Hence, the expression in (7.41) is positive and condition (7.37)(i) is met.

To complete the proof of convexity of each  $\theta_t(\tilde{h}_t, \tilde{h}_{t+1})$  function and hence the proof of the convexity of  $U(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t])$  in the  $\tilde{h}_t$  variables, it must be shown that condition (7.37)(ii) is satisfied by the cubic utility function in (7.30). Substituting from (7.38), (7.39), and (7.40) into (7.37)(ii), one obtains

$$(7.42) \quad \left\{ -W_t \frac{u_t''(W_t)}{[u_t'(W_t)]^3} \right\} \left\{ -W_{t+1} \frac{u_{t+1}''(W_{t+1})}{[u_{t+1}'(W_{t+1})]^3} \right\} - \left\{ \frac{1}{[u_t'(W_t)]^2} \right\} \left\{ \frac{1}{[u_{t+1}'(W_{t+1})]^2} \right\} > 0 .$$

Clearly, the condition in (7.42) will be fulfilled, for all  $t=1, \dots, T-1$  if

$$(7.43) \quad -W_t \frac{u_t''(W_t)}{[u_t'(W_t)]^3} > \frac{1}{[u_t'(W_t)]^2} \quad \text{for each } t=1, \dots, T.$$

The condition in (7.43) is equivalent to

$$(7.44) \quad -\frac{1}{[u_t'(W_t)]^3} \{ W_t u_t''(W_t) + u_t'(W_t) \} > 0.$$

Since  $u_t'(W_t) > 0$ , the condition in (7.42) will be fulfilled if the term in braces in (7.44) is negative for each  $t$ . Substituting the appropriate expressions for  $u_t'(W_t)$  and  $u_t''(W_t)$  one finds that the term in braces will be negative if and only if

$$(7.45) \quad 9W_t^2 - 8k_t W_t + k_t^2 + g_t^2 < 0.$$

This is equivalent to

$$(7.46) \quad (W_t - \frac{4}{9} k_t)^2 - \frac{1}{9} (\frac{7}{9} k_t^2 - g_t^2) < 0.$$

The restriction in (7.30) on  $g_t^2$  and  $k_t^2$  ensures that  $\frac{7}{9} k_t^2 - g_t^2 > 0$  so that  $-\frac{1}{9} (\frac{7}{9} k_t^2 - g_t^2) < 0$ . In addition, the restriction in (7.30) that  $W_t$  lie between

$\frac{4}{9} k_t - \frac{1}{9} \sqrt{7k_t^2 - 9g_t^2}$  and  $\frac{4}{9} k_t + \frac{1}{9} \sqrt{7k_t^2 - 9g_t^2}$  guarantees that

$(W_t - \frac{4}{9} k_t)^2 < \frac{1}{9} (\frac{7}{9} k_t^2 - g_t^2)$ . Thus, the condition in (7.46), and therefore the one

in (7.37)(ii), is fulfilled by the utility function (7.30) for every  $t$ . The

proof of the convexity of  $U(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W_t^t, t])$  in the  $\tilde{h}_t$  variables is complete

and condition (7.9)(i) is fulfilled.

One fact ought to be noted before going on to proving that the function in Theorem 7.2 satisfies (7.9)(ii). Specifically, the utility function in (7.30) would be nonexistent if it were impossible to meet the three restrictions on the

ranges of the  $W_t$ 's simultaneously with parameters such that  $\frac{1}{3} k_t^2 < g_t^2 < \frac{7}{9} k_t^2$ . That is, one wants to be sure that there exist  $g_t^2, k_t^2$  pairs such that the intersection of  $W_t > 0$ ,

$$(7.47) \quad \left[ \frac{2}{3} k_t - \frac{1}{3} \sqrt{3g_t^2 - k_t^2}, \frac{2}{3} k_t \right) \quad \text{and}$$

$$(7.48) \quad \left( \frac{4}{9} k_t - \frac{1}{9} \sqrt{7k_t^2 - 9g_t^2}, \frac{4}{9} k_t + \frac{1}{9} \sqrt{7k_t^2 - 9g_t^2} \right)$$

is nonempty when  $\frac{1}{3} k_t^2 < g_t^2 < \frac{7}{9} k_t^2$ . In short, the minimum of the two lower bounds in (7.47) and (7.48) must be positive, the maximum of them must be smaller than the minimum of the two upper bounds, and simultaneously the condition  $\frac{1}{3} k_t^2 < g_t^2 < \frac{7}{9} k_t^2$  must obtain.

The condition that  $W_t > 0$  is equivalent to the simultaneous fulfillment of

$$(7.49) \quad \frac{2}{3} k_t - \frac{1}{3} \sqrt{3g_t^2 - k_t^2} > 0 \quad \text{and} \quad \frac{4}{9} k_t - \frac{1}{9} \sqrt{7k_t^2 - 9g_t^2} > 0.$$

But,

$$(7.50) \quad \frac{2}{3} k_t - \frac{1}{3} \sqrt{3g_t^2 - k_t^2} = k_t \left( \frac{2}{3} - \frac{1}{3} \sqrt{3 \frac{g_t^2}{k_t^2} - 1} \right) \quad \text{and}$$

$$(7.51) \quad \frac{4}{9} k_t - \frac{1}{9} \sqrt{7k_t^2 - 9g_t^2} = k_t \left( \frac{4}{9} - \frac{1}{9} \sqrt{7 - 9 \frac{g_t^2}{k_t^2}} \right).$$

Clearly, the expression in (7.50) will be smallest for  $\frac{g_t^2}{k_t^2}$  largest, while the expression in (7.51) will be minimal for  $\frac{g_t^2}{k_t^2}$  smallest. Given that  $\frac{1}{3} < \frac{g_t^2}{k_t^2} < \frac{7}{9}$ , one sees that

$$(7.52) \quad \frac{2}{3} k_t - \frac{1}{3} \sqrt{3g_t^2 - k_t^2} > \frac{2}{3} k_t \left( 1 - \frac{1}{\sqrt{3}} \right) > 0 \quad \text{for } k_t > 0, \text{ and}$$

$$(7.53) \quad \frac{4}{9} k_t - \frac{1}{9} \sqrt{7k_t^2 - 9g_t^2} > \frac{2}{9} k_t > 0 \quad \text{for } k_t > 0 .$$

Hence, given the restriction  $\frac{1}{3} k_t^2 < g_t^2 < \frac{7}{9} k_t^2$ , the minimum value for  $W_t$  as allowed by the intervals in (7.30) is always positive.

As for the nonemptiness of the intersection of the intervals in (7.47) and (7.48) when the restriction on  $g_t^2$  and  $k_t^2$  holds, it is clear that each lower bound in (7.47) and (7.48) is always smaller than its corresponding upper bound. In addition, it is also obvious that the lower bound in (7.48) is always smaller than the upper one in (7.47). It remains to be shown that there exist  $g_t^2, k_t^2$  pairs with  $k_t > 0$  such that

$$(7.54) \quad \frac{2}{3} k_t - \frac{1}{3} \sqrt{3g_t^2 - k_t^2} < \frac{4}{9} k_t + \frac{1}{9} \sqrt{7k_t^2 - 9g_t^2} \quad \text{and} \quad \frac{1}{3} < \frac{g_t^2}{k_t^2} < \frac{7}{9}$$

hold simultaneously. But the first inequality in (7.54) is equivalent to

$$(7.55) \quad \frac{2}{9} < \frac{1}{3} \sqrt{3 \frac{g_t^2}{k_t^2} - 1} + \frac{1}{9} \sqrt{7 - 9 \frac{g_t^2}{k_t^2}} .$$

The right-hand side is a function of  $\frac{g_t^2}{k_t^2}$ , the first and second derivatives of which are

$$(7.56) \quad \frac{1}{2} \left( \frac{1}{\sqrt{3 \frac{g_t^2}{k_t^2} - 1}} - \frac{1}{\sqrt{7 - 9 \frac{g_t^2}{k_t^2}}} \right) \text{ and}$$

$$(7.57) \quad -\frac{1}{4} \left\{ \frac{3}{\left(3 \frac{g_t^2}{k_t^2} - 1\right)^{\frac{3}{2}}} + \frac{9}{\left(7 - 9 \frac{g_t^2}{k_t^2}\right)^{\frac{3}{2}}} \right\} , \text{ respectively.}$$

But from the second derivative in (7.57), it is clear that the function is strictly concave for all  $\frac{g_t^2}{k_t^2}$  for which it assumes real values, since it has been assumed throughout that when a square root is taken, it is always the principal or positive value one takes. In addition, from (7.56) it can be seen that when  $\frac{g_t^2}{k_t^2} = \frac{1}{3}$  the expression on the right-hand side of (7.55) is increasing since its first derivative then equals  $+\infty$ . When, at the other extreme,  $\frac{g_t^2}{k_t^2} = \frac{7}{9}$ , the expression on the right-hand side of (7.55) is decreasing since its first derivative then equals  $-\infty$ .

In short,  $\frac{1}{3} \sqrt{3 \frac{g_t^2}{k_t^2} - 1} + \frac{1}{9} \sqrt{7 - 9 \frac{g_t^2}{k_t^2}}$  is a continuous strictly concave function of  $\frac{g_t^2}{k_t^2}$  that attains its maximum value in the interval  $(\frac{1}{3}, \frac{7}{9})$ . Hence, its minimum value over the interval  $\frac{1}{3} < \frac{g_t^2}{k_t^2} < \frac{7}{9}$  must be assumed at one of the endpoints of the interval. Thus, over the interval  $\frac{1}{3} < \frac{g_t^2}{k_t^2} < \frac{7}{9}$ , the expression on the right-hand side of (7.55) is greater than

$$\min \left\{ \frac{1}{3} \sqrt{3 \cdot \frac{1}{3} - 1} + \frac{1}{9} \sqrt{7 - 9 \cdot \frac{1}{3}}, \frac{1}{3} \sqrt{3 \cdot \frac{7}{9} - 1} + \frac{1}{9} \sqrt{7 - 9 \cdot \frac{7}{9}} \right\} = \frac{2}{9} .$$

But then the inequality in (7.55) obtains for all  $g_t^2, k_t^2$  pairs satisfying the restrictions in (7.30). Thus, the ranges of  $W_t$  specified in (7.30) and repeated in (7.47) and (7.48) have points in common for all  $g_t^2, k_t^2$  pairs satisfying the restrictions in (7.30) and  $W_t$  is always positive for any  $g_t^2, k_t^2$  pair satisfying those restrictions.

$$(ii) \quad dU \left( \sum_{t=1}^T U^{-1}[\tilde{h}_t, W_t^t, t] \right) = dU \left( \sum_{t=1}^T [W_t^t + \tilde{Z}^t] \right) = dU(W + \tilde{Z}),$$

where the differential is with respect to nonnegative increments in the initial consumption incomes, and where we have used (7.16) and the definitions of  $W^t$  and  $\tilde{Z}^t$ . Therefore,

$$(7.58) \quad dU\left(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t]\right) = \sum_{t=1}^T \left(\frac{\partial U}{\partial W_t}\right)_{W+Z} \tilde{d}W_t.$$

But from the result in (7.31) for the marginal utility of each  $W_t$ , this is equivalent to

$$(7.59) \quad \left\{ \begin{aligned} dU\left(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t]\right) &= \sum_{t=1}^T [3(W_t + \tilde{z}_t)^2 - 4k_t(W_t + \tilde{z}_t) + k_t^2 + g_t^2] dW_t \\ &+ m_1(W_2 + \tilde{z}_2) dW_1 + \sum_{t=2}^{T-1} [m_{t-1}(W_{t-1} + \tilde{z}_{t-1}) + m_t(W_{t+1} + \tilde{z}_{t+1})] dW_t \\ &+ m_{T-1}(W_{T-1} + \tilde{z}_{T-1}) dW_T. \end{aligned} \right.$$

The sum of convex functions is, however, convex. In discussing decreasing risk aversion for the additive cubic utility function, it was proven that the first sum on the right-hand side of (7.59) is convex in the  $\tilde{h}_t$  variables, since that sum is the expression in (7.20), which was shown to be convex. Hence, the total differential  $dU\left(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W^t, t]\right)$  will be convex in the  $\tilde{h}_t$  variables if the remaining sum on the right-hand side of (7.59) is convex.

That remaining sum is a nonnegative linear combination of the  $W_t + \tilde{z}_t$  variables because each  $m_t > 0$  and each  $dW_t \geq 0$ . Hence, from the discussion of equation (7.17) it is known that the sum of those remaining terms in (7.59) is a nonnegative linear combination of the  $u_t^{-1}(\tilde{h}_t)$  functions. The expression for  $\frac{d^2 u_t^{-1}(\tilde{h}_t)}{d\tilde{h}_t^2}$  in (7.39) combined with the fact that  $u_t'(W_t) > 0$  for all  $W_t$  and  $u_t''(W_t) < 0$ , due to the restrictions on the range of  $W_t$ , imply that

$\frac{d^2 u_t^{-1}(\tilde{h}_t)}{d\tilde{h}_t^2}$  is positive. Each  $u_t^{-1}(\tilde{h}_t)$  function is, therefore, convex in its respective  $\tilde{h}_t$ . Since a nonnegative linear combination of convex functions is convex, the sum of the terms that appear in (7.59) in addition to those from (7.20) is convex, as desired.

Thus,  $dU(\sum_{t=1}^T U^{-1}[\tilde{h}_t, W_t^t, t])$  is a sum of convex functions in the  $\tilde{h}_t$  variables and is therefore itself a convex function of them. Condition (7.9)(ii) is satisfied by the utility function of Theorem 7.2.

(iii) Setting all  $W_i = 0$  except  $W_t$ ,  $\frac{\partial^3 U}{\partial W_t^3} \cdot \frac{\partial U}{\partial W_t} \geq (\frac{\partial^2 U}{\partial W_t^2})^2$ , and this must be true for all  $t$ . Setting all  $W_i = 0$  except  $W_t$  in (7.30) yields  $u_t(W_t)$ , just as it did with the separable function considered earlier. But then the condition just stated in (iii) is equivalent to that in (7.26) and (7.27), which is true for the proposed utility function as a result<sup>25</sup> of the restrictions

$$g_t^2 > \frac{1}{3} k_t^2 \quad \text{and} \quad \frac{2}{3} k_t - \frac{1}{3} \sqrt{3g_t^2 - k_t^2} \leq W_t < \frac{2}{3} k_t$$

This completes the proof that the utility function defined in (7.30) subject to the restrictions on its parameters and on the ranges of consumption incomes stated in (7.30) satisfies the set of sufficient conditions (7.6)-(7.10). It therefore satisfies the set of axioms presented earlier for a utility function that is to serve as the capital-budgeting problem's objective function.

#### 7.4.D. A Possible Interpretation of the Proposed Utility Function

Returning from this lengthy digression with the assurance that a multiperiod cubic utility function can comprehend interperiod interactions between consumption incomes, attention from now on will be focused on the additive cubic utility function presented in (7.11). As the proof of

<sup>25</sup>The proof is exactly the same as the proof of the validity of (7.26) in the proof for the separable cubic utility function.

Theorem 7.1 showed, that utility function satisfies all the axioms it was stipulated the basis for a capital-budgeting objective function should satisfy. One possible interpretation of the additive cubic function in (7.11) is the following. A decisionmaker possessing the attitudes we have set out as axioms of rationality appears on the scene. He specifies the ranges of consumption incomes he believes relevant and over which he possesses the views described. Then, in accordance with (7.11), restrictions on the  $k_t$  and  $g_t$  parameters can be specified so that a cubic of the form proposed can describe his attitudes toward risk over these ranges.

Note that the validity of the cubic describing his behavior need not be restricted to the particular relevant ranges he specifies. That is, the cubic finally chosen by him as best describing his views when more specific considerations than positive marginal utility, risk aversion, and decreasing risk aversion in the multiperiod sense are taken into account may display these risk-behavior conditions over ranges wider than the ones relevant to him. But, the cubic of the type in (7.11) finally settled upon must meet these requirements for at least his relevant ranges of consumption incomes.

It is not being asserted that the functions in (7.11) and (7.30) are the only ones that can express the decisionmaker's attitudes toward risk over the ranges of consumption incomes he thinks relevant. They are not the only functions that can satisfy the axioms. In fact, as a re-examination of the proof of Theorem 7.1 will show, any additive utility function in which each single period's component shows positive marginal utility, is bounded, risk-averse, and single-period decreasingly risk-averse could describe such a decisionmaker's attitudes. For example, a utility function that sums appropriately bounded logarithmic utility functions,  $\log(W_t + \delta_t)$  with  $\delta_t > 0$  and defined for  $W_t \leq W_t^*$ , could also express the decisionmaker's risk attitudes.<sup>26</sup> The logarithmic function does

---

<sup>26</sup>See Pratt (1964, p. 133) for examples of other single-period decreasingly risk-averse utility functions. Note, however, that some of the functions he lists are unbounded until one restricts the range of relevant consumption income. He does not present the single-period cubic discussed here.



not, however, meet the fifth property stipulated for a utility function. It would require full specification of the density function of the  $W_t$ 's. What Theorem 7.1 and the comments preceding it state is that the additive cubic utility function in (7.11) is the lowest degree polynomial utility function that meets all the desiderata we have specified.<sup>27</sup>

7.5. The Capital-Budgeting Utility Function and Attitudes Toward Risk Characteristics

With the postulated additive cubic serving as the decisionmaker's multiperiod utility function, consider how the expected-utility maximizer behaves. For expositional convenience let us change the notation in that cubic utility function. Letting  $K_t = 2k_t$  and  $G_t = k_t^2 + g_t^2$ , one can write

$$(7.60) \quad u_t(W_t) = W_t^3 - K_t W_t^2 + G_t W_t \quad \text{and}$$

$$(7.61) \quad U(W) = \sum_{t=1}^T u_t(W_t) = \sum_{t=1}^T [W_t^3 - K_t W_t^2 + G_t W_t] .$$

The function is, of course, again descriptive of the decisionmaker's attitudes only for the ranges in (7.11). The individual maximizing expected utility with (7.61) as his utility function will be maximizing

$$(7.62) \quad \sum_{t=1}^T [E(u_t)] = \sum_{t=1}^T [E(W_t^3) - K_t E(W_t^2) + G_t E(W_t)] .$$

---

<sup>27</sup>The author has been able to locate one previous case in the literature where a cubic utility function is suggested for decisionmaking under risk. In discussing the reinsurance problem of an insurance company, Borch (1962, pp. 183-188) introduces the single-period cubic utility function  $U(W) = (W-a)^3 + b$ . For one thing, as he recognizes, this function has increasing marginal utility for all  $W > a$ . Hence, for consumption income in excess of  $a$  the decisionmaker is a risk-lover. This is not damaging for one can restrict  $W < a$ , as was done in (7.11). But what is problematic about Borch's proposal is that it is everywhere increasingly risk-averse. It cannot meet the axiom of decreasing risk aversion over any range of consumption income.

Examine more closely the tth component of this sum. As for notation, let  $\mu_{W_t}$  be the mean of  $W_t$ ,  $\sigma_{W_t}^2$  the variance of  $W_t$ , and  $\phi_{W_t}$  the third moment of  $W_t$  about its mean, that is,  $\phi_{W_t} = E(W_t - \mu_{W_t})^3$ . For any random variable  $X$ , the following relationships obtain

$$(7.63) \quad \begin{cases} E(X^2) = \sigma_X^2 + \mu_X^2 & \text{and} \\ E(X^3) = \phi_X + 3\mu_X\sigma_X^2 + \mu_X^3 \end{cases} .$$

The third moment about the mean is often used as a measure of skewness (or asymmetry), either by itself or after having been normalized in terms of units of measurement.<sup>28</sup> Symmetric distributions always have zero values for  $\phi_X$ . Distributions with large right-hand tails will normally have large positive values for  $\phi_X$  while those with large left-hand tails will generally have large (in absolute value) negative values for  $\phi_X$ . Unfortunately, asymmetric distributions may sometimes yield a zero value for  $\phi_X$ .<sup>29</sup> In general, the weight of opinion is that the third moment about the mean can serve as a weak measure of the skewness of the distribution.

Examining the tth component of the sum in (7.62), substitute for  $E(W_t^3)$  and  $E(W_t^2)$  using the expressions in (7.63). One finds, after rearranging terms, that,

$$(7.64) \quad E(u_t) = \phi_{W_t} + (3\mu_{W_t} - K_t) \sigma_{W_t}^2 + \mu_{W_t}^3 - K_t \mu_{W_t}^2 + G_t \mu_{W_t} .$$

The individual optimizing the utility function in (7.11), or (7.61), under risk acts only on the basis of mean, variance, and skewness. (The third moment about the mean is henceforth identified with skewness.) His attitudes toward each of

<sup>28</sup>For a sampling of the statistical literature on the matter see Cramér (1946, pp. 183-184); Hoel (1962, p. 77); Kendall and Stuart (1963, p. 85); and Mood and Graybill (1963, pp. 109-110).

<sup>29</sup>See Hoel (1962, p. 77) and Mood and Graybill (1963, pp. 109-110).

these properties of a consumption income's probability distribution can be easily ascertained.

First, the decisionmaker prefers ceteris paribus increases in the expected value of the distribution of income. To see this, consider the change in  $E(u_t)$  resulting from a ceteris paribus change in  $\mu_{W_t}$ . One finds

$$(7.65) \quad \frac{\partial E(u_t)}{\partial \mu_{W_t}} = 3\sigma_{W_t}^2 + 3\mu_{W_t}^2 - 4k_t \mu_{W_t} + (k_t^2 + g_t^2) \quad ,$$

after substituting for  $K_t$  and  $G_t$ . But the discussion of (7.13) and (7.14), the marginal utilities of the different periods' incomes, indicates that  $3\mu_{W_t}^2 - 4k_t \mu_{W_t} + k_t^2 + g_t^2$  will be positive for all  $\mu_{W_t}$ . Moreover,  $3\sigma_{W_t}^2$  will always be nonnegative. Hence the right-hand side of (7.65) is positive for all  $\mu_{W_t}$ , as asserted. Given the variance and skewness of a distribution, the decisionmaker with utility function (7.61) prefers a distribution with a higher mean to one with a lower mean.

Second, the owner of the utility function in (7.61) shuns ceteris paribus increases in variance. Differentiating (7.64) partially with respect to  $\sigma_{W_t}^2$  and substituting  $2k_t$  for  $K_t$  yields

$$(7.66) \quad \frac{\partial E(u_t)}{\partial (\sigma_{W_t}^2)} = 3\mu_{W_t} - 2k_t \quad .$$

But the bounds on  $W_t$  in (7.11) ensure that  $W_t < \frac{2}{3} k_t$ . If every possible relevant consumption income in period  $t$  is less than  $\frac{2}{3} k_t$ , then the mean of  $W_t$  must be less than  $\frac{2}{3} k_t$ . Hence  $3\mu_{W_t} - 2k_t < 0$  and, as stated, the expression on the right-hand side of (7.66) is negative. Given the mean and skewness of a distribution of consumption income, that distribution has a higher expected utility the lower is its variance.

The results about ceteris paribus changes in the mean and ceteris paribus changes in the variance of the distribution of consumption income in period  $t$  are not different from those found in the literature.<sup>30</sup> The next attitude of the decisionmaker is, however, not commonly found in writings on decisionmaking under risk. Accepting the third moment as an indication of skewness, the decisionmaker's expected utility is increased by ceteris paribus increments in positive skewness. This is clear because

$$(7.67) \quad \frac{\partial E(u_t)}{\partial \phi_{W_t}} = 1 > 0 .$$

Given the mean and variance of the distribution of consumption income, the decisionmaker with (7.11) or (7.61) as his utility function will prefer a distribution more skewed to the right than one less skewed to the right. For a given expected value and a given degree of dispersion of outcomes, the decisionmaker prefers a greater chance of extreme favorable outcomes and generally moderate losses to a greater chance of extreme unfavorable outcomes and generally moderate gains.

To avoid any misunderstandings and to dispel any suspicions about this preference for positive (rightward) skewness, it is important to note that this preference is fundamental in terms of the axiom set presented. Each of the single-period utility functions,  $u_t(W_t)$ , that sum to yield  $U(W)$  must display decreasing risk aversion. This, it will be recalled, requires that

$$(7.26') \quad u_t'(W_t)u_t'''(W_t) \geq [u_t''(W_t)]^2 .$$

But then certainly  $u_t'(W_t)u_t'''(W_t)$  must be positive. With each period's consumption income required by the axioms to have positive marginal utility, we must have  $u_t'(W_t) > 0$ . Hence, it follows that  $u_t'''(W_t)$  must be positive which means that

---

<sup>30</sup>Recall the previous discussion of the mean-variance approach. See, for example, Markowitz (1959, pp. 139-144).

the sign of the cubic term must be positive if the utility function is only a third-degree polynomial. In the case of a cubic utility function, however, the sign of the partial derivative of  $E(u_t)$  with respect to  $\phi_{W_t}$  is simply the sign of the cubic term. Thus, the partial derivative in (7.67) must be positive as a result of the axioms stipulated.

To summarize, the decisionmaker whose multiperiod utility function is the additive cubic in (7.11), which was shown to satisfy the axioms, derives added expected utility from ceteris paribus increases in mean and skewness, respectively, and from ceteris paribus decreases in variance. It is very interesting to note that in a recent empirical study of utility functions of middle-management executives, P. E. Green actually observed such tradeoffs between the moments of the distribution of returns. A preference for smaller variance and for positive skewness was exhibited. When a distribution had a greater variance than a given control distribution, a higher expected return than the control's was required; and when it had a greater degree of positive (negative) skewness but the same variance as the control the decisionmaker was content (only content) with a lower (higher) expected return than that offered by the control.<sup>31</sup>

To conclude this section, note that in writing about the shortcomings of the mean-variance approach, K. Borch states that

The obvious limitation of the method is that only the two first [sic] moments are considered when the ordering is established over the set of probability distributions. There seems to be good reason for bringing in the third moment, and thus assuming that the skewness of the probability distributions is considered in decision-making under uncertainty. This will mean that the utility function  $u(x)$  can be polynomial of third degree, and this class of functions agrees fairly well with our intuitive ideas of the 'utility of money.'<sup>32</sup>

---

<sup>31</sup>Green (1963, pp. 39-40, especially Figure 5).

<sup>32</sup>Borch (1963, p. 700).

In the development just completed, the necessity for bringing in the third moment, considering the skewness of a distribution, and using a polynomial of at least the third degree as the utility function has been derived from a set of axioms for a suitable utility function. This development has provided Borch with his "good reason" and has provided a suitable objective function for capital budgeting under risk.

#### 7.6. A Programming Model for Capital Budgeting Under Risk

The goal of this chapter was to develop a programming model for capital-budgeting decisions in a specific risk environment. The nature of that risk environment, as it was described in Chapter 6, ensures that the constraint set faced by the firm is the same under risk as it was under certainty. It only remains, then, to express the objective function already developed and stated in (7.62) -- that is, maximizing the expected value of the utility function in (7.11) -- in terms of the projects, the elements of choice in the model.

##### 7.6.A. Some Additional Assumptions

At this point the set of assumptions that have been built into the model for reasons of tractability and expositional convenience is enlarged. First, we anticipate the inability of presently available methods, and the decomposition method to be used in the next chapter in particular, to cope with an objective function containing: (i) a nonlinear form of integer-valued variables plus (ii) nonlinear forms of individual continuous variables -- specifically, second and third powers of each of them -- plus (iii) the sum of products of the first and second powers of each of the continuous variables with a different nonlinear form of integer-valued variables (each such form also differing from the one in (i)). The nonlinear forms of the integer-valued variables can be linearized by generalizing to higher-order interactions one of the methods for treating cash-flow interactions discussed in Section 5.5.

This does not, however, remove the great analytical and computational difficulties involved in (a) the simultaneous presence of a linear form in integer variables and second and third powers of continuous variables, and (b) the existence of a sum of products of linear forms in the integer variables with the first and second powers of the continuous variables. But it is precisely this type of objective function that one obtains when

$$(7.68) \quad W_t = \sum_{i=1}^n a_{ti} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n a_{tij} y_i y_j + (1+r_L)R_t$$

is substituted for  $W_t$  in the utility function in (7.61) and the expectations operator is applied.

This difficulty will be avoided by assuming that  $R_t = 0$  for all  $t$ . This must appear quite strong for apparently now the firm cannot set aside for disbursement any of the cash throw-offs from its previously owned resources. The situation is not that bad. All that is required is that some of the projects considered by the firm be disbursement activities.

A disbursement project for the  $t$ th period, for example, would entail a demand of a fixed sum on the funds available in the  $t$ th period and would yield a deterministic return of  $1+r_L$  times that fixed sum at the end of that period. There could be several alternative "cash-removal" projects available to the firm in any given period. They would correspond to the withdrawal of alternative amounts of the previous cash throw-off from the firm's originally owned resources. The cash-removal projects in a period would not be mutually exclusive, except as constrained by the size of the previous cash throw-offs, so that a number of levels of withdrawal would be possible in any one period. It is not clear that this alternative approach to disbursement of funds is as overly restrictive as it might first appear to be. "Real-world" corporations do,

after all, often choose their particular disbursements to owners on the basis of a yes-no decision with respect to a fixed "normal" amount, with special additional outpayments ("extra dividends") occurring in packages of fixed size.

Treating the cash-removal activity of the firm as a discrete project -- or set of discrete projects -- in each period has one important implication for the constraint set of the programming model. It was argued in Chapter 5 that the continuous nature of the lending and withdrawal activities implied that at the optimum each of the firm's budget constraints would actually hold as an equation, not as an inequality. With lending for present distribution or for future use providing a return, none of the firm's funds would ever rest idle. The corporation would either move funds that were not needed in period  $t$  to a later period when they would be needed or it would disburse those funds with the added benefit of the one-period return to lending. With the cash-removal activities discrete, however, the budget constraints no longer need hold as equations in the optimal solution. Whether or not they will appear as equations at the optimum will depend on the sizes of the cash-withdrawal activities the firm makes available to itself in the several periods.

Clearly, the firm could always lend any extra funds it might have available in a period after it had provided for (1) the physical projects' cash requirements in that period, (2) any debt repayment in that period, (3) any movement of funds to later periods that is required if the planned investment program is to be undertaken, and (4) any planned cash withdrawal in that period. The point is simply that due to the discrete nature of the disbursement activities as they now appear in the model, further lending or carrying over of funds to future periods may be fruitless. The funds resulting from such lending may not be large enough to enable undertaking an additional or a larger disbursement activity in any later period. In such a case, while the firm could lend the excess money, it would be indifferent to doing so since its objective-function



value would be unaffected by such lending. If it does lend this unusable excess, the constraint holds as an equation. If it does not lend the excess (and ineffectual) funds, the particular constraint holds as an inequality. Hence, while all of the firm's budget constraints can still hold as equations at the optimum, it is no longer necessarily true that they will.

The second assumption to be added at this time is simply that for expositional convenience attention will continue to be confined to interactions between two projects. The method used for coping with these interactions can easily be extended to higher-order interactions. But there is little, if anything, to be gained in terms of analytical insights and much to be lost in expositional clarity by including interactions among more than two projects. Little enough attention has been paid to second-order interrelationships in much of the previous capital-budgeting literature so that by treating them fully we are taking, I think, enough of a step forward at this time.

Interdependences of higher than the second order occur in the present model in the form of joint product moments involving three or more individual projects. These higher-order interrelationships may take the form of product moments involving individual-project cash flows or cash-flow interactions or both. Examples of such joint product moments -- where  $i, j, k, \ell, p,$  and  $q$  are distinct projects -- are:  $E(a_{ti}a_{tj}a_{tk})$ ,  $E(a_{tij}a_{tk})$ ,  $E(a_{tij}^2a_{tk})$ ,  $E(a_{tij}a_{tk\ell})$ , and  $E(a_{tij}a_{tk\ell}a_{tpq})$ . It is assumed, from now on, that any product moment about the origin that would involve project interactions among three or more projects (up to six in the present model) is zero.

One final assumption will be made for expositional purposes. In order to ease notational problems, it will also be assumed that the product moment about the origin of any nonzero power of a cash-inflow interaction with any nonzero power of either of the interacting projects is also zero. That is, it is now assumed that  $E(a_{tij}^{\eta_1} a_{tk}^{\eta_2}) = 0$  whenever  $k$  equals either  $i$  or  $j$  and  $\eta_1 + \eta_2$  equals

two or three with  $\eta_1 \geq 0$  and  $\eta_2 \geq 0$ . Hence, combining the assumption in the previous paragraph and the present one, in the present model,  $E(a_{tij}^{\eta_1} a_{tk}^{\eta_2}) = 0$  for any  $i, j, k$  triplet with  $i < j$ ,  $\eta_1 + \eta_2$  equal to two or to three, and  $\eta_1 \geq 0$  and  $\eta_2 \geq 0$ .

### 7.6.B. The Objective Function Expressed in Terms of the Projects

With these simplifying assumptions in hand, consider expressing the objective function

$$(7.62) \quad \text{Maximize } E\{U(W)\} = \sum_{t=1}^T E\{u_t(W_t)\} = \sum_{t=1}^T [E(W_t^3) - K_t E(W_t^2) + G_t E(W_t)]$$

in terms of the individual projects of the model. Notationally, let  $\mu_{ti}$  denote the mean of the probability distribution of  $a_{ti}$ ,  $\mu_{tii}$  is that distribution's second moment about the origin, and  $\mu_{tiii}$  that distribution's third moment about the origin. The symbol  $\mu_{tij}$  denotes the product moment (about the origin) of the distributions of  $a_{ti}$  and  $a_{tj}$ . That is,  $\mu_{tij}$  equals  $E(a_{ti} a_{tj})$  not  $E(a_{tij})$ . Proliferating subscripts once again, the product moment about the origin  $E(a_{ti}^2 a_{tj})$  shall be written as  $\mu_{tiji}$ . Finally, the first, second, and third moments of  $a_{tij}$  about the origin will be written as  $E(a_{tij})$ ,  $E(a_{tij}^2)$  and  $E(a_{tij}^3)$  respectively.

The definition of  $W_t$  with  $R_t$  set equal to zero is

$$(7.69) \quad W_t = \sum_{i=1}^n a_{ti} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n a_{tij} y_i y_j$$

Using this definition and the assumptions about joint product moments made in Section 7.6.A one obtains the following expressions.

$$(7.70) \quad \mu_{W_t} = E(W_t) = \sum_{i=1}^n \mu_{ti} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n E(a_{tij}) y_i y_j$$

$$(7.71) \quad \mu_{W_t^2} = E(W_t^2) = \sum_{i=1}^n \mu_{tiii} y_i + 2 \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n \mu_{tij} y_i y_j + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n E(a_{tij}^2) y_i y_j .$$

$$(7.72) \quad \mu_{W_t^3} = E(W_t^3) = \sum_{i=1}^n \mu_{tiii} y_i + 3 \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n (\mu_{tiii} + \mu_{tijj}) y_i y_j + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n (a_{tij}^3) y_i y_j .$$

In writing (7.71) and (7.72) the fact that, with each  $y_i$  equal to zero or one,  $y_i = y_i^2 = y_i^3$  has been used. Grouping terms in  $y_i$  and  $y_i y_j$  and substituting the expressions in (7.70), (7.71), and (7.72) into the  $t$ th component of (7.62) one obtains

$$(7.73) \quad \left\{ \begin{aligned} E\{u_t(W_t)\} &= \sum_{i=1}^n (\mu_{tiii} - K_t \mu_{tii} + G_t \mu_{ti}) y_i \\ &+ \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n [3\mu_{tiii} + 3\mu_{tijj} + E(a_{tij}^3) - 2K_t \mu_{tij} - K_t E(a_{tij}^2) + G_t E(a_{tij})] y_i y_j . \end{aligned} \right.$$

The stochastic interdependences among projects that are induced by the presence of risk should be clear. These stochastic interactions appear because of the nonlinear form of the utility function. A nonlinear utility function, it will be recalled, is definitely necessitated by the presence of risk, although I have argued for the need for such nonlinearity even under certainty. The interrelationships between projects that will appear in the capital-budgeting objective function resulting from summing the expression in (7.73) over all  $T$  periods are thus of two types: the risk-induced ones in the form of the joint product moments of  $a_{ti}$ 's and  $a_{tj}$ 's for  $i \neq j$  and the nonstochastic ones in the form of the moments of the direct cash-flow interactions, the  $a_{tij}$ 's.

The artificial-project approach to capturing cash-flow interactions presented for use in the certainty model can clearly also be used in the case.

of the stochastic interdependences between projects present in the objective function with (7.73) as its tth component. Introduce one incremental artificial-project variable,  $x_{ij}$ , for each  $i, j$  pair with  $i < j$ , and subject  $x_{ij}$  to the constraints

$$(7.74) \quad \left\{ \begin{array}{l} -y_i + x_{ij} \leq 0 \\ -y_j + x_{ij} \leq 0 \\ y_i + y_j - x_{ij} \leq 1 \\ x_{ij} \geq 0 \end{array} \right. .$$

Using these artificial projects to rewrite the expression in (7.73) and summing the resulting expression for  $E\{u_t(W_t)\}$  over the  $T$  periods of the model, the objective function for capital budgeting under risk is obtained. It is

$$(7.75) \quad \begin{aligned} \text{Maximize} \quad & \sum_{i=1}^n \left[ \sum_{t=1}^T \{ \mu_{tiii} - K_t \mu_{tii} + G_t \mu_{ti} \} \right] y_i \\ & + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n \left[ \sum_{t=1}^T \{ 3\mu_{tiij} + 3\mu_{tijj} + E(a_{tij}^3) - 2K_t \mu_{tij} - K_t E(a_{tij}^2) + G_t E(a_{tij}) \} \right] x_{ij} . \end{aligned}$$

The programming model of capital budgeting under risk is now complete. The decisionmaker's objective function is (7.75), which he maximizes subject to the constraint set in (7.76).

### 7.7. A Recapitulation and A Look Ahead

This programming model resulted from a reconsideration of the certainty model of Chapter 5, when the corporation faced the risk environment described in Section 6.2 rather than a deterministic world. Shortcomings of previous proposals, both practical and theoretical, for budgeting capital in such a risk environment led to a programming model and to a set of desiderata for an objective function

$$\sum_{i=1}^n c_{li} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n c_{lij} x_{ij} + \ell_1 - b_1 \leq C_0 .$$

$$\sum_{i=1}^n c_{ti} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n c_{tij} x_{ij} - (1+r_L) \ell_{t-1} + \ell_t + (1+r_B) b_{t-1} - b_t \leq C_{t-1}$$

for  $t=2, \dots, T-1$ .

$$\sum_{i=1}^n c_{Ti} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n c_{Tij} x_{ij} - (1+r_L) \ell_{T-1} + (1+r_B) b_{T-1} \leq C_{T-1} .$$

$$\sum_{i=1}^n d_{ki} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n d_{kij} x_{ij} \leq D_k \quad \text{for } k=1, \dots, K .$$

$$b_t \leq B_t \quad \text{for } t=1, \dots, T-1 .$$

(7.76)  $\sum_{i \in S} y_i \leq 1$  for each set S of mutually exclusive projects.

$$y_i - y_j \leq 0 \quad \text{for each contingent project (i)-independent project (j) pair .}$$

$$\left. \begin{aligned} -y_i + x_{ij} &\leq 0 \\ -y_j + x_{ij} &\leq 0 \\ y_i + y_j - x_{ij} &\leq 1 \end{aligned} \right\} \text{for each } i, j \text{ pair with } i < j .$$

$$y_i = 0 \text{ or } 1 \text{ for each } i .$$

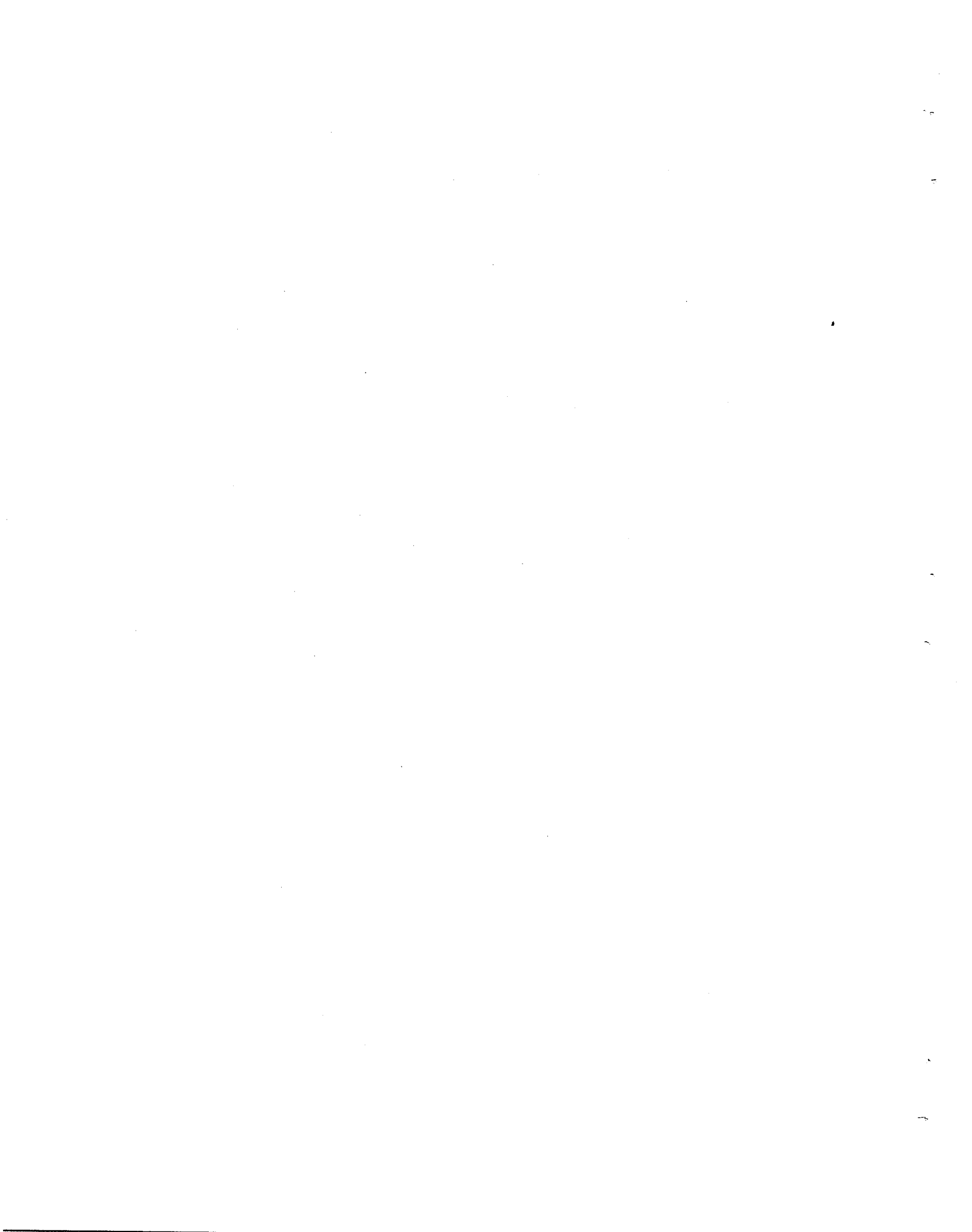
$$x_{ij} \geq 0 \text{ for each } i, j \text{ pair with } i < j .$$

$$b_t \geq 0 \text{ for each } t=1, \dots, T-1 .$$

$$\ell_t \geq 0 \text{ for each } t=1, \dots, T-1 .$$

for that model. With the expected-utility theorem serving as the basis for rational behavior under risk, a cubic utility function was proven to be the lowest-degree polynomial utility function satisfying the set of axioms developed. The possibility of incorporating interactions between different periods' consumption incomes into the determination of multiperiod utility was demonstrated. But for the sake of expositional simplicity the presentation was pursued in terms of an additive cubic utility function. The decisionmaker with such a utility function was shown to prefer, ceteris paribus, higher values of expected return to lower values of expected return, smaller variances to larger variances, and a greater degree of positive skewness. Finally, after making several further assumptions, in the interest of analytical tractability and expositional effectiveness, about the firm and the environment in which it operates, the objective function and the constraint set that constitute the programming model of capital budgeting under risk were presented.

The task of the next chapter will be to show how the programming model with (7.75) as the objective function and (7.76) as the constraint set can be solved. A partitioning algorithm due to J. F. Benders will be used to solve the problem. As shall be seen, an interesting economic interpretation can be lent to the algorithm's progress toward a solution.



## CHAPTER 8

### SOLVING THE PROBLEM OF CAPITAL BUDGETING UNDER RISK

#### 8.1. Introduction

The discussion in preceding chapters has argued that the correct approach to capital budgeting under risk is one of mathematical programming. In particular, the linear mixed-integer program with which the last chapter closed has been presented as the appropriate model for capital budgeting given the risk environment and imperfect capital market described earlier. A decisionmaker is, however, interested in more than correctly formulating the problem he faces. He wants to know how to solve his problem within the framework of that correct formulation. The present study has proceeded with this goal of the decisionmaker in the forefront, as attested to by the assumptions made for the sake of analytical tractability. It is to the question of actually solving the problem of capital budgeting under risk as formulated in this study that the present chapter turns.

There exists, at present, a number of methods specifically developed for solving linear mixed-integer programming problems.<sup>1</sup> The solution procedure presented here is based upon a decomposition algorithm due to J. F. Benders.<sup>2</sup> The chapter begins with a description of Benders' partitioning procedure for a general linear mixed-integer programming problem. Then the application of Benders' method to the problem of capital budgeting under risk is discussed.

---

<sup>1</sup>See, for example, Beale (1958), Beale and Small (1965), Driebeek (1966), Gomory (1960b), Harris (1964), and Land and Doig (1960). A concise presentation of several of these alternatives appears in the excellent survey article by M. L. Balinski (1965).

<sup>2</sup>Benders (1962). Benders' method is applicable to mixed-variable problems more general in form than the linear mixed-integer programming problem. The more general problem it can be used to solve will be indicated in footnote 4.



The algorithm's progress toward a solution of the capital-budgeting problem can be lent an interesting economic interpretation. This interpretation is presented next. The chapter continues with a discussion of some particular aspects of the solution procedure and their economic meaning. Lastly, a summary of the iterative decision procedure is presented.

In the course of what follows, the reader's indulgence is requested as his ability to be flexible with notation may be tried (perhaps severely) at some points. In some cases, new variables or parameters may have symbolic representations similar to those of other parameters or variables that have appeared in previous chapters. When this occurs, the new meaning attached to the symbol is what it is intended to represent from the point of its introduction onward.

### 8.2. Benders' Decomposition Applied to a Maximization Linear Mixed-Integer Programming Problem<sup>3</sup>

The general linear mixed-integer programming problem, of which the capital-budgeting model developed here is a special case, takes the form

$$\begin{aligned} & \text{Maximize} \quad Px + Qy \\ (8.1) \quad & \text{Subject to: } A_1x + A_2y \leq B \\ & \quad \quad \quad x \geq 0 \\ & \quad \quad \quad y \in R. \end{aligned}$$

In the problem in (8.1),  $A_1$  and  $A_2$  are matrices while  $x$ ,  $y$ ,  $P$ ,  $Q$ , and  $B$  are all vectors. Since the problem in (8.1) is a linear mixed-integer program, the set  $R$ , in which the vector  $y$  is constrained to lie, is the set of all vectors (of the dimension of  $y$ ) with integral elements.<sup>4</sup>

<sup>3</sup>The description to be presented is an adaptation of the presentations in Balinski and Wolfe (1963, pp. 3-7) and Balinski (1965, pp. 271-274). These articles describe the application of Benders' procedure to a minimization linear mixed-integer problem.

<sup>4</sup>Benders' procedure is applicable to problems of the more general form

Following the lead of the Balinski-Wolfe exposition, Benders' method seems to be seen best through the following equivalence transformations of the original problem in (8.1). That program may be written as

$$(8.2) \quad \text{Max } \{ Px + Qy \mid A_1x + A_2y \leq B, x \geq 0, y \in R \} ,$$

or as the iterated maximization problem

$$(8.3) \quad \text{Max}_{y \in R} \{ Qy + \max_x [ Px \mid A_1x \leq B - A_2y, x \geq 0 ] \} .$$

A hint as to how Benders' method partitions the problem can be gleaned from (8.3). Those variables restricted to some subset (R) of the reals are separated from the ones that are only constrained to be nonnegative and to satisfy the restraints  $A_1x + A_2y \leq B$ . Two programming problems -- one in the former set of variables and one in the latter -- are then solved in sequence.

In order to add substance to this vague notion of partitioning, note that for a given y-vector, the maximization problem within the curly braces is a linear program in x. This is true because Px is a linear form in x, and with y given the constraints  $A_1x \leq B - A_2y$  and  $x \geq 0$  constitute a set of linear constraints. But then there exists a dual minimization linear program

$$(8.4) \quad \text{Min}_z \{ z(B - A_2y) \mid zA_1 \geq P, z \geq 0 \}$$

where z is the vector of dual variables.<sup>5</sup> The duality theory of linear programming then provides the result

---

$$\begin{aligned} &\text{Maximize} && Px + f(y) \\ &\text{Subject to:} && A_1x + f(y) \leq B \\ &&& x \geq 0 \\ &&& y \in R . \end{aligned}$$

That is, it can be used when the variables (the elements of y) that are restricted to some arbitrary subset of the real numbers appear in nonlinear expressions in the objective function and in the constraints. See Benders (1962). The discussion here is confined to the more restricted case where all variables appear only in linear forms and where R is the set of all vectors with only integer components.

<sup>5</sup>This is one instance of a second use of a previously employed symbol. The symbol z from now on denotes the vector of dual variables and has no relation to the vectors of risks discussed in earlier chapters.

$$(8.5) \quad \text{Max}_x \{ Px | A_1 x \leq B - A_2 y, x \geq 0 \} = \text{Min}_z \{ z(B - A_2 y) | zA_1 \geq P, z \geq 0 \} .$$

The program on the left-hand side of (8.5) has "value"  $+\infty$  if it is unbounded, and the minimization problem on the right-hand side has "value"  $-\infty$  if it is unbounded. The original problem, (8.1), is therefore equivalent to

$$(8.6) \quad \text{Max}_{y \in R} \{ Qy + \min_z [z(B - A_2 y) | zA_1 \geq P, z \geq 0] \} .$$

The set of vectors  $\Omega = \{z | zA_1 \geq P, z \geq 0\}$  is a convex polyhedral set in the space with dimension equal to the dimension of vector  $z$ . Note that  $\Omega$  is independent of the particular values assumed by the elements of the vector  $y$ . If the set  $\Omega$  is empty, then clearly the minimization problem in (8.4) and (8.5) is infeasible. It follows from duality theory that the maximization linear program in  $x$  alone, in (8.5), then has no solution. It is either infeasible or it has an unbounded objective function. Hence, if  $\Omega$  is empty, the original mixed-integer problem has no solution.

On the other hand, if  $\Omega$  is not empty, the minimum value of  $z(B - A_2 y)$ , for any  $y$ , is achieved at an extreme point of  $\Omega$  or  $z(B - A_2 y)$  is unbounded along an extreme ray of  $\Omega$  since  $\Omega$  is a convex polyhedral set. The extreme points of  $\Omega$  are, however, finite in number (say,  $M$ ) and the extreme rays of  $\Omega$  are also finite in number (say,  $H$ ). Label the extreme points  $z^m$ ,  $m \in \bar{M}$ , and the extreme rays  $z^h$ ,  $h \in \bar{H}$ , where  $\bar{M} = \{1, 2, \dots, M\}$  and  $\bar{H} = \{1, 2, \dots, H\}$ .

Consider the extreme rays of the convex polyhedral set  $\{z | zA_1 \geq P, z \geq 0\}$ . They are found by locating the extreme rays of the finite convex cone  $\{z | zA_1 \geq 0, z \geq 0\}$ . If for some  $y$  there exists an extreme ray  $z^h$  with  $z^h(B - A_2 y) < 0$ , the minimization problem in (8.4) will clearly have an unbounded objective function. The common value of the two linear programs in (8.5) will then be  $-\infty$  and the maximization problem in (8.5) will be

infeasible. But, if for the given  $y$ ,  $z^h(B-A_2y) \geq 0$  for all  $h \in \bar{H}$ , the minimization program in (8.4) has a finite solution. The constraint set  $A_1x \leq B-A_2y$ ,  $x \geq 0$ , then admits a feasible solution for  $x$ . Necessary and sufficient conditions on  $y$  so that it admits a feasible solution to the maximization linear program in (8.5) are thus

$$(8.7) \quad z^h(B-A_2y) \geq 0 \quad \text{for all } h \in \bar{H} .$$

The mixed-integer program with which we began, (8.1), can consequently be written as the following all-integer program in  $y$  alone:

$$(8.8) \quad \begin{array}{l} \text{Max } \{\Psi \mid \Psi \leq Qy + \text{Min}_{m \in \bar{M}} z^m(B-A_2y) \text{ , and } z^h(B-A_2y) \geq 0 \text{ for all } h \in \bar{H} \} . \\ y \in R \end{array}$$

The original problem could then be solved in two steps. First, the all-integer program in (8.8) would be solved. If it is found to be infeasible, then the original problem to which it is equivalent is infeasible. If (8.8) is found to be feasible but have no (optimal) solution -- that is,  $\Psi$  unbounded -- then problem (8.1) also lacks an optimal solution. The third alternative is that the problem in (8.8) is feasible with an optimal solution. The solution of the pure-integer programming problem in (8.8) would then yield an optimal vector  $y^0$  and an optimal solution value  $\Psi^0$  for the original mixed-integer program. The second step of the partitioning procedure would then be to use the optimal vector  $y^0$  to solve the maximization linear program in (8.5) for the optimum  $x$ -vector,  $x^0$ . The solution to the original problem, (8.1), is then complete: optimum value =  $\Psi^0$ ,  $x = x^0$ , and  $y = y^0$ .

The number of constraints in (8.8) is finite because the number of extreme points and the number of extreme rays of  $\Omega$  are finite. Finite may, however, be very large. Hence, one may be confronted with the undesirable and practically impossible task of enumerating a very large number of linear constraints on  $y$  and  $\Psi$ . The hope behind Benders' computational procedure is

that only a small number of extreme points of  $\Omega$  and only a small number of extreme rays of  $\Omega$  need ever be used to generate constraints. Our concern, after all, is not with the constraint set in (8.8) but with the optimal solution to the mixed-integer program in (8.1). The algorithm proceeds through a sequence of steps, then, which leads to the generation of a set of constraints -- based on a subset of the union of  $\bar{M}$  and  $\bar{H}$  -- that determines the optimum solution of the original problem.

Each step of the procedure involves the solution of two subproblems: one is a general programming problem (in the present case an integer program) and the second is a linear programming problem.

(a) Given a finite set of  $z^s$ ,  $s \in S$  where  $S \subset \bar{M} \cup \bar{H}$  (that is,  $S$  is some subset of the union of the sets of extreme points and extreme rays of  $\Omega$ ), solve the restricted integer programming problem (8.8) over the set of constraints deriving from the elements of  $S$ . That is, solve

$$(8.9) \quad \begin{array}{l} \text{Max } \{ \Psi \leq Qy + \text{Min}_{s \in S \cap \bar{M}} z^s (B-A_2 y) \text{ and } z^s (B-A_2 y) \geq 0 \text{ for } s \in S \cap \bar{H} \} \\ y \in R \end{array}$$

If the problem is infeasible, so is the original problem in (8.1) since the constraints based on the set  $S$  make up only a subset of the total number of constraints based on  $\bar{M} \cup \bar{H}$ . If the problem is feasible, let  $(\bar{\Psi}, \bar{y})$  be the optimal solution pair for (8.9). (If  $\Psi$  is unbounded above in (8.9), take  $\bar{\Psi}$  large and  $\bar{y}$  feasible.)

(b) To determine whether or not this solution  $(\bar{\Psi}, \bar{y})$  is optimal for (8.1), solve the linear programs in (8.5) with  $y = \bar{y}$ . Let their common value be  $f(\bar{y})$ . If  $f(\bar{y}) = -\infty$ , the maximization linear program is infeasible: the current  $\bar{y}$  does not admit a feasible  $x$ . With  $f(\bar{y}) = -\infty$ , the minimization linear program has an unbounded objective function and a new extreme ray  $z^h$ ,  $h \in \bar{H}$  has been found as the "solution" vector to that minimization program in (8.5). It is adjoined to the set  $S$  and we return to step (a).

If  $f(\bar{y}) = +\infty$ , then the minimization linear program in (8.4) and (8.5) is infeasible. That is, the set  $\Omega$  is empty and there is no solution to the original mixed-integer program.

If  $f(\bar{y})$  is finite, suppose  $\bar{x}$  and  $\bar{z}$  are the solutions to the respective linear programs in (8.5). Clearly,  $(\bar{x}, \bar{y})$  is a feasible solution vector for the original problem (8.1). If, in addition,  $\bar{\Psi} \leq Q\bar{y} + f(\bar{y}) = Q\bar{y} + P\bar{x}$ , then  $(\bar{x}, \bar{y})$  is also optimal for the mixed-integer problem (8.1). This is true because  $(\bar{\Psi}, \bar{y})$  is feasible for the complete integer program in (8.8) while  $\Psi$  cannot be any larger than its maximum value in the restricted problem (8.9) solved in part (a). That is, the maximum value  $\Psi$  can assume for the entire program (8.1) is limited by the constraints imposed on it in (8.8). Since there are more constraints in (8.8) than in the restricted integer program (8.9), the maximum value  $\Psi$  can assume in (8.8) is less than or equal to the maximum value it can attain in (8.9). But having found a feasible pair for (8.1), namely,  $(\bar{x}, \bar{y})$ , with a value for  $\Psi$  for the entire program that is greater than or equal to this maximum of  $\Psi$  in the restricted program (8.9), the optimum for the entire problem (8.1) or (8.8) has been found. It is  $(\bar{x}, \bar{y})$ .

If, on the other hand,  $\bar{\Psi} > Q\bar{y} + f(\bar{y})$ , some linear constraint(s) in (8.8) is (are) violated by the solution  $(\bar{\Psi}, \bar{y})$  to the restricted integer program (8.9). The one "most violated" is the one corresponding to  $\bar{z}$ , that is,  $\Psi \leq Qy + \bar{z}(B-A_2y)$ . Hence,  $\bar{z}$  is adjoined to the set  $S$  and we return to step (a).

After each step, then, exactly one of several alternatives has occurred. We may have found that the original problem has no solution. This can be discovered in one of two ways. First, we may find that the set  $\{z | zA_1 \geq P, z \geq 0\}$  is empty. Since this set is independent of  $y$ , we will find out whether or not

it is empty at the first performance of step (b). Second, we may have found that the original mixed-integer program (8.1) -- or its equivalent (8.8) -- is infeasible because we find that the restricted pure integer program in (8.9) is infeasible. Third, we may have found the optimum for the given programming problem (8.1), namely,  $(\bar{\Psi}, \bar{x}, \bar{y})$ , where  $(\bar{\Psi}, \bar{y})$  solves (8.9) and  $\bar{x}$  solves the resulting maximization linear program in (8.5). Fourth, we may have found a new extreme ray to add to the set S, which occurs if  $f(\bar{y}) = -\infty$ . Or, finally, we may have found a new extreme point to add to the set S, which occurs if  $f(\bar{y})$  is finite but  $\bar{\Psi} > Q\bar{y} + f(\bar{y})$ .

The procedure terminates within a finite number of steps, either with the conclusion that the original problem is infeasible or that it is feasible but has no finite optimal solution, or with the optimal solution to the problem. The finiteness of the procedure is guaranteed because (i) the set  $\Omega$  has only a finite number of extreme rays and a finite number of extreme points, and (ii) at each step at which the procedure does not terminate a new one of these rays or points is found. The aim and the hope of Benders' procedure is to determine the optimum to the mixed-integer program with the aid of only a small number of these extreme rays and extreme points. The hope of the present author is that Benders' method is especially successful in solving the problem of capital budgeting under risk. Let us return to that problem now.

### 8.3. Benders' Decomposition and the Capital-Budgeting Model

The model of capital budgeting under risk developed thus far is summarized in the mixed-integer program consisting of (7.75) and (7.76). For convenience, that programming problem is repeated here, with one change for notational and expositional simplicity. Let  $Q_i = \sum_{t=1}^T (\mu_{tiii} - K_t \mu_{tii} + G_t \mu_{ti})$ , which is the ith project's individual contribution to expected utility, and

$$P_{ij} = \sum_{t=1}^T \{3\mu_{t|ij} + 3\mu_{t|jj} + E(a_{tij}^3) - 2K_t \mu_{tij} - K_t E(a_{tij}^2) + G_t E(a_{tij})\},$$

which is the joint contribution of projects  $i$  and  $j$  to the expected utility of the investment program. The model of capital budgeting under risk to which Benders' partitioning procedure will be applied is the mixed-integer programming problem in (8.10).

The program in (8.10) is clearly a special case of the general mixed-integer problem presented in (8.1) for which the application of Benders' procedure has just been described. The continuous-valued variables are the artificial projects ( $x_{ij}$ 's), the lending activities ( $l_t$ 's), and the borrowing activities ( $b_t$ 's). The variables restricted to an arbitrary subset of the reals are the  $y_i$ -variables, corresponding to the real physical projects and fixed-size disbursement activities, which must be zero or one. That is,  $R$  in (8.10) is the set of all  $n$ -element vectors whose elements are zeros or ones.

An interesting aspect of the mixed-integer program in (8.10) is the existence of linear constraints in which either only continuous-valued variables appear or only zero-one variables appear. The former are the set of borrowing-limits constraints; the latter are the constraints expressing mutual-exclusion and contingency relationships among projects. In terms of the general mixed-integer program in (8.1), there are rows of  $A_1$  that have all zero elements (the mutual-exclusion and contingency constraints) and there are rows of  $A_2$  that have all elements equal to zero (the borrowing-limits constraints).

At the outset let us note that the program in (8.10) is feasible. With each  $C_{t-1}$ ,  $D_k$ , and  $B_t$  nonnegative, the null vector ( $y_i = x_{ij} = b_t = l_t = 0$  for all  $i, j, t$ ) provides a feasible solution. In essence, the firm always has the alternative of doing nothing and, hence, the mixed-integer programming model of the capital-budgeting problem always has a feasible solution. This means that



$$\text{Maximize } \sum_{i=1}^n Q_i y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n P_{ij} x_{ij}$$

Subject to:

$$\sum_{i=1}^n c_{li} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n c_{lij} x_{ij} + l_1 - b_1 \leq C_0 .$$

$$\sum_{i=1}^n c_{ti} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n c_{tij} x_{ij} - (1+r_L)l_{t-1} + l_t + (1+r_B)b_{t-1} - b_t \leq C_{t-1}$$

for  $t=2, \dots, T-1$ .

$$\sum_{i=1}^n c_{Ti} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n c_{Tij} x_{ij} - (1+r_L)l_{T-1} + (1+r_B)b_{T-1} \leq C_{T-1} .$$

$$\sum_{i=1}^n d_{ki} y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n d_{kij} x_{ij} \leq D_k \quad \text{for } k=1, \dots, K .$$

(8.10)  $\left\{ \begin{array}{l} b_t \leq B_t \quad \text{for } t=1, \dots, T-1 . \end{array} \right.$

$$\sum_{i \in S} y_i \leq 1 \quad \text{for each set } S \text{ of mutually exclusive projects.}$$

$$y_i - y_j \leq 0 \quad \text{for each contingent project (i)-independent project (j) pair.}$$

$$\left. \begin{array}{l} -y_i + x_{ij} \leq 0 \\ -y_j + x_{ij} \leq 0 \\ y_i + y_j - x_{ij} \leq 1 \end{array} \right\} \text{ for each } i, j \text{ pair with } i < j .$$

$$y_i = 0 \quad \text{or } 1 \text{ for each } i .$$

$$x_{ij} \geq 0 \quad \text{for each } i, j \text{ pair with } i < j .$$

$$b_t \geq 0 \quad \text{for each } t=1, \dots, T-1 .$$

$$l_t \geq 0 \quad \text{for each } t=1, \dots, T-1 .$$

when Benders' procedure is used, a solution  $(\bar{\Psi}, \bar{y})$  is always obtained in part (a) of each Benders step. Note, in addition, that the capital-budgeting problem in (8.10) always has a finite objective-function value since each  $Q_i$ ,  $P_{ij}$ ,  $y_i$ , and  $x_{ij}$  is finite. This finiteness property and the feasibility of (8.10) have an important implication about the linear programs appearing in part (b) of a Benders step, which are shown in (8.11) and (8.12) and will be discussed in more detail presently. The dual linear program (8.12) must always be feasible because if the dual were infeasible the original problem in (8.10) would have to be either infeasible or unbounded, neither of which is possible. Specifically, if the maximization problem in (8.11) is infeasible, the dual problem in (8.12) must be unbounded since it cannot possibly be infeasible.

$$\begin{aligned}
 & \text{Maximize} && \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n P_{ij} x_{ij} \\
 & \text{Subject to:} && \\
 & \text{(i)} & \left\{ \begin{aligned} & \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n c_{lij} x_{ij} + l_1 - b_1 \leq C_0 - \sum_{i=1}^n c_{li} \bar{y}_i . \\ & \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n c_{tij} x_{ij} - (1+r_L) l_{t-1} + l_t + (1+r_B) b_{t-1} - b_t \leq C_{t-1} - \sum_{i=1}^n c_{ti} \bar{y}_i . \\ & \hspace{15em} \text{for } t=2, \dots, T-1. \\ & \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n c_{Tij} x_{ij} - (1+r_L) l_{T-1} + (1+r_B) b_{T-1} \leq C_{T-1} - \sum_{i=1}^n c_{Ti} \bar{y}_i . \end{aligned} \right. \\
 & \text{(ii)} & \left\{ \begin{aligned} & \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n d_{kij} x_{ij} \leq D_k - \sum_{i=1}^n d_{ki} \bar{y}_i \quad \text{for } k=1, \dots, K . \end{aligned} \right. \\
 & \text{(iii)} & b_t \leq B_t \quad \text{for } t=1, \dots, T-1 . \\
 & \text{(iv)} & x_{ij} \leq \bar{y}_i \\
 & \text{(v)} & x_{ij} \leq \bar{y}_j \\
 & \text{(vi)} & -x_{ij} \leq 1 - \bar{y}_i - \bar{y}_j \\
 & \text{(vii)} & x_{ij} \geq 0 \quad \text{for each } i, j \text{ pair with } i < j . \\
 & \text{(viii)} & b_t \geq 0, \quad l_t \geq 0 \quad \text{for each } t=1, \dots, T-1 .
 \end{aligned}$$

$$\text{Minimize } \sum_{t=1}^T [C_{t-1} - \sum_{i=1}^n c_{ti} \bar{y}_i] \rho_t + \sum_{k=1}^K [D_k - \sum_{i=1}^n d_{ki} \bar{y}_i] \lambda_k + \sum_{t=1}^{T-1} B_t \theta_t$$

$$+ \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n \{ \bar{y}_i u_{ij} + \bar{y}_j v_{ij} + (1 - \bar{y}_i - \bar{y}_j) w_{ij} \}$$

Subject to:

(8.12) (i)  $\sum_{t=1}^T c_{tij} \rho_t + \sum_{k=1}^K d_{kij} \lambda_k + u_{ij} + v_{ij} - w_{ij} \geq P_{ij}$

for each  $i, j$  pair with  $i < j$  .

(ii)  $\rho_t - (1+r_L) \rho_{t+1} \geq 0$  for all  $t=1, \dots, T-1$  .

(iii)  $-\rho_t + (1+r_B) \rho_{t+1} + \theta_t \geq 0$  for all  $t=1, \dots, T-1$  .

(iv)  $\rho_t \geq 0$  for all  $t=1, \dots, T$ ;  $\lambda_k \geq 0$  for all  $k=1, \dots, K$ ;

$\theta_t \geq 0$  for all  $t=1, \dots, T-1$  .

(v)  $u_{ij} \geq 0$  ,  $v_{ij} \geq 0$  ,  $w_{ij} \geq 0$  for each  $i, j$  pair with  $i < j$  .

### 8.3.A. The Linear Subproblem

Continuing with the application of Benders' procedure to the capital-budgeting model in (8.10), let us consider the linear subproblem in more detail. Corresponding to the dual linear programming problems in (8.5), for given values of the zero-one (real-project and disbursement-activity) variables  $\bar{y} = \{\bar{y}_i\}$  and the corresponding value of  $\Psi$  , there is the maximization problem in (8.11) and the minimization program in (8.12). In the minimization program,  $\rho_t$  is the dual variable corresponding to the  $t$ th budget constraint in (8.11),  $\lambda_k$  is the dual variable corresponding to the  $k$ th limited-resource constraint in (8.11), and  $\theta_t$  is the dual variable corresponding to the  $t$ th borrowing limit in (8.11). The variables

$u_{ij}$ ,  $v_{ij}$ , and  $w_{ij}$  are the duals corresponding to the constraints of types (iv) (v), and (vi), respectively, on  $x_{ij}$  in the primal linear program.

In some applications of Benders' partitioning procedure, it is convenient to enumerate the extreme rays of  $\Omega$  at the outset and to determine then what conditions can be imposed on  $y$  to ensure that every solution to the restricted integer program in (8.9) admits a feasible  $x$  to the maximization linear program in (8.5).<sup>6</sup> Unfortunately, this is not the case with the problem at hand. A search for conditions on  $y = \{y_i\}$  that guarantee the existence of a feasible solution  $\{\{x_{ij}\}, \{b_t\}, \{\theta_t\}\}$  to the maximization program in (8.11) consists of looking for necessary and sufficient restrictions on  $y$  such that

$$\begin{aligned}
 & \left. \begin{aligned}
 (i) \quad & \sum_{t=1}^T [C_{t-1} - \sum_{i=1}^n c_{ti} \bar{y}_i] \rho_t + \sum_{k=1}^K [D_k - \sum_{i=1}^n d_{ki} \bar{y}_i] \lambda_k + \sum_{t=1}^{T-1} B_t \theta_t \\
 & + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n \{ \bar{y}_i u_{ij} + \bar{y}_j v_{ij} + (1 - \bar{y}_i - \bar{y}_j) w_{ij} \} \geq 0 \\
 & \text{for all extreme rays } \{ \{ \rho_t \}, \{ \lambda_k \}, \{ \theta_t \}, \{ u_{ij} \}, \{ v_{ij} \}, \{ w_{ij} \} \} \text{ of} \\
 (8.13) \quad & \left. \begin{aligned}
 (ii) \quad & \sum_{t=1}^T c_{tij} \rho_t + \sum_{k=1}^K d_{kij} \lambda_k + u_{ij} + v_{ij} - w_{ij} \geq 0 \text{ for each } i, j \text{ pair with } i < j . \\
 (iii) \quad & \rho_t - (1+r_L) \rho_{t+1} \geq 0 \text{ for all } t=1, \dots, T-1 . \\
 (iv) \quad & -\rho_t + (1+r_B) \rho_{t+1} + \theta_t \geq 0 \text{ for all } t=1, \dots, T-1 . \\
 (v) \quad & \{ \{ \rho_t \}, \{ \lambda_k \}, \{ \theta_t \}, \{ u_{ij} \}, \{ v_{ij} \}, \{ w_{ij} \} \} \geq 0 .
 \end{aligned}
 \right\}
 \end{aligned}
 \right\}
 \end{aligned}$$

In the present model it is best to introduce extreme-ray constraints as they arise -- as described in the earlier presentation of Benders' algorithm -- just as one introduces new extreme-point constraints as they emerge from the solution of the linear subproblem.

<sup>6</sup>See Balinski (1965, p. 273) for this suggestion and Balinski and Wolfe (1963) and Balinski (1965, pp. 289-290) for an example of its use.

Given a candidate capital budget, that is, a vector  $\bar{y} = \{\bar{y}_i\}$  of zeros and ones, the primal linear program in (8.11) involves no optimization. The  $\bar{y}_i$ -values and the constraints (iv)-(vii) on each  $x_{ij}$ -variable together imply a unique set of values for the  $x_{ij}$ -variables. The value of the optimal solution to (8.11) -- indeed, the only feasible value of the linear program in (8.11) if it has a feasible solution vector -- is then immediately obtained by using this set of  $x_{ij}$ -values to

$$\text{compute } \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n P_{ij} x_{ij} .$$

The only question with regard to the maximization linear program is whether the given  $\bar{y}$  admits a feasible solution. The question really being asked is: Does there exist a feasible pattern of borrowing and lending -- where the borrowing limits and the last-period budget constraint are the binding factors -- and do there exist adequate amounts of the other resources ( $D_k$ 's) to enable the firm to pursue the capital-investment program given by  $\bar{y}$ ? If there does exist such a schedule of borrowing and lending and if there do exist adequate resources to render the candidate budget feasible, then  $f(\bar{y})$  is given by  $\sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n P_{ij} \bar{x}_{ij}$ ,

where  $\{\bar{x}_{ij}\}$  is the vector of  $x_{ij}$ -values implied by the candidate budget.

The question of the feasibility of the linear program in (8.11) given  $\bar{y} = \{\bar{y}_i\}$  is easily resolved by the following forward-substitution process.

- (a) Find the  $x_{ij}$ -values implied by the  $\bar{y}$ -vector using constraints (iv)-(vii) in (8.11). Denote them as  $\bar{x}_{ij}$ , for each  $i, j$  with  $i < j$ .
- (b) Substitute these  $\bar{x}_{ij}$ -values into the limited-resource constraints, (ii) in (8.11). If these constraints are not satisfied,  $\bar{y}$  does not admit a feasible solution to (8.11). If these constraints are satisfied, go on to (c).
- (c) Substitute the  $\bar{x}_{ij}$ -values into the first-period budget constraint.

From that first-period constraint one then obtains a constraint of the form  $l_1 - b_1 \leq \phi_1$ , where  $\phi_1$  is a constant.

(1) If  $\phi_1 \geq 0$ , set  $\bar{l}_1 = \phi_1$ ,  $\bar{b}_1 = 0$ . Go on to the second-period budget constraint.

(2) If  $\phi_1 < 0$ , set  $\bar{l}_1 = 0$ ,  $\bar{b}_1 = -\phi_1$ . Check the first-period borrowing limit. If  $\bar{b}_1 > B_1$ ,  $\bar{y}$  does not admit a feasible solution to (8.11). If  $\bar{b}_1 \leq B_1$ , go on to the second-period budget constraint.

(d) Considering the tth-period budget constraint for  $t=2, \dots, T-1$ , substitute the  $\bar{x}_{ij}$ -values and the just-derived  $\bar{l}_{t-1}$ ,  $\bar{b}_{t-1}$ -values into the constraint. The tth-period budget constraint then yields a constraint of the form  $\bar{l}_t - b_t \leq \phi_t$ , where  $\phi_t$  is a constant.

(1) If  $\phi_t \geq 0$ , set  $\bar{l}_t = \phi_t$ ,  $\bar{b}_t = 0$ . Go on to the t+1st-period constraint.

(2) If  $\phi_t < 0$ , set  $\bar{l}_t = 0$ ,  $\bar{b}_t = -\phi_t$ . Check the tth-period borrowing limit. If  $\bar{b}_t > B_t$ ,  $\bar{y}$  does not admit a feasible solution to (8.11). If  $\bar{b}_t \leq B_t$ , go on to the t+1st-period budget constraint.

(e) For the Tth-period budget constraint, substituting the  $\bar{x}_{ij}$ ,  $\bar{l}_{T-1}$ , and  $\bar{b}_{T-1}$  values into the constraint, one obtains a value

$$\phi_T = C_{T-1} - \sum_{i=1}^n c_{ti} \bar{y}_i - \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n c_{tij} \bar{x}_{ij} + (1+r_L) \bar{l}_{T-1} - (1+r_B) \bar{b}_{T-1} .$$

If  $\phi_T < 0$ ,  $\bar{y}$  does not admit a feasible solution to the linear program in (8.11). If  $\phi_T \geq 0$ ,  $\bar{y}$  does admit a feasible solution to that program, namely, the feasible solution  $\{ \{\bar{x}_{ij}\}, \{\bar{b}_t\}, \{\bar{l}_t\} \}$  that has just been found by construction.

There are, then, two possibilities. Either the candidate investment program  $\bar{y}$  does not admit a feasible solution to the linear program in (8.11) or it does admit a feasible solution to that problem. Considering these in turn, suppose first that  $\bar{y}$  does not admit a feasible solution to the maximization linear program. The dual linear program in (8.12) then has an unbounded objective

function. The minimum value of the objective function in (8.12) goes to  $-\infty$  along an extreme ray of the convex polyhedral cone  $\Gamma$  defined by (ii)-(v) of (8.13). The "solution" to the dual linear program, that is, the extreme ray of  $\Gamma$  along which the objective function goes to  $-\infty$ , generates a constraint that must be added to the restricted integer subproblem. Specifically, suppose this is the  $h$ th extreme ray enumerated. The constraint that must be added is

$$(8.14) \quad \sum_{t=1}^T [C_{t-1} - \sum_{i=1}^n c_{ti} y_i] \rho_t^h + \sum_{k=1}^K [D_k - \sum_{i=1}^n d_{ki} y_i] \lambda_k^h + \sum_{t=1}^{T-1} B_t \theta_t^h + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n \{y_i u_{ij}^h + y_j v_{ij}^h + (1-y_i-y_j) w_{ij}^h\} \geq 0 .$$

The candidate budget  $\bar{y}$  does not satisfy this constraint because it is with  $y = \bar{y}$  that the objective function in the minimization problem goes to  $-\infty$  along the ray  $\{ \{\rho_t^h\}, \{\lambda_k^h\}, \{\theta_t^h\}, \{u_{ij}^h\}, \{v_{ij}^h\}, \{w_{ij}^h\} \}$ . Hence, the constraint in (8.14) serves to remove from future consideration the proposed investment program given by  $\bar{y} = \{\bar{y}_i\}$ . Constraints of this type thus serve to exclude as possibilities proposed budgets which are, in fact, infeasible, when project interactions are taken into account, because of financial constraints (the borrowing limits or  $T$ th-period budget effecting the infeasibility) or because of constraints on other resources of the firm.

Turning to the second possible result when  $\bar{y}$  is substituted into the linear programs in (8.11) and (8.12), consider the case in which  $\bar{y}$  does admit a feasible  $\{ \{\bar{x}_{ij}\}, \{\bar{b}_t\}, \{\bar{\ell}_t\} \}$ -vector to the primal maximization problem. Then,

$$\begin{aligned}
 f(\bar{y}) = & \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n P_{ij} \bar{x}_{ij} = \sum_{t=1}^T [C_{t-1} - \sum_{i=1}^n c_{ti} \bar{y}_i] \bar{\rho}_t + \sum_{k=1}^K [D_k - \sum_{i=1}^n d_{ki} \bar{y}_i] \bar{\lambda}_k \\
 (8.15) \quad & + \sum_{t=1}^{T-1} B_t \bar{\theta}_t + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n [\bar{y}_i \bar{u}_{ij} + \bar{y}_j \bar{v}_{ij} + (1 - \bar{y}_i - \bar{y}_j) \bar{w}_{ij}]
 \end{aligned}$$

is most easily calculated using the first equation in (8.15), that is, the  $P_{ij}$ - and  $\bar{x}_{ij}$ -values. If it is found that

$$(8.16) \quad \bar{\Psi} \leq \sum_{i=1}^n Q_i \bar{y}_i + f(\bar{y}) = \sum_{i=1}^n Q_i \bar{y}_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n P_{ij} \bar{x}_{ij} ,$$

then  $\bar{y} = \{\bar{y}_i\}$  is the optimal capital budget, and the capital-budgeting problem under risk is solved.

Suppose, on the other hand, that

$$(8.17) \quad \bar{\Psi} > \sum_{i=1}^n Q_i \bar{y}_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n P_{ij} \bar{x}_{ij} .$$

In this case a new extreme point of the set defined by the constraints of the dual linear program (8.12) has been found. The constraint corresponding to this extreme point must be added to the restricted integer program of the type in (8.9). If this is the mth extreme point enumerated, the constraint that must be added is

$$\begin{aligned}
 \Psi \leq & \sum_{i=1}^n Q_i y_i + \sum_{t=1}^T [C_{t-1} - \sum_{i=1}^n c_{ti} y_i] \rho_t^m + \sum_{k=1}^K [D_k - \sum_{i=1}^n d_{ki} y_i] \lambda_k^m + \sum_{t=1}^{T-1} B_t \theta_t^m \\
 (8.18) \quad & + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n [y_i u_{ij}^m + y_j v_{ij}^m + (1 - y_i - y_j) w_{ij}^m] .
 \end{aligned}$$



8.3.B. The Restricted Integer Subproblem

The constraints of the type in (8.18) have an interesting economic interpretation. Before discussing it, however, consider the restricted integer subproblem of the capital-budgeting model in which these constraints play an important role. The restricted integer subproblem for the present model begins with a given set of extreme points and a given set of extreme rays of the convex polyhedral set defined by the constraints in (8.12). Suppose up to the present step  $H^0$  extreme rays and  $M^0$  extreme points have been enumerated, where  $H^0 < H$  and  $M^0 < M$ . The restricted integer program is then

Maximize  $\Psi$

Subject to:

(i) 
$$\Psi \leq \sum_{i=1}^n Q_i y_i + \sum_{t=1}^T [C_{t-1} - \sum_{i=1}^n c_{ti} y_i] \rho_t^m + \sum_{k=1}^K [D_k - \sum_{i=1}^n d_{ki} y_i] \lambda_k^m + \sum_{t=1}^{T-1} B_t \theta_t^m$$

$$+ \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n [y_i u_{ij}^m + y_j v_{ij}^m + (1-y_i-y_j)w_{ij}^m]$$

for  $m=1, \dots, M^0$  ;

(ii) 
$$0 \leq \sum_{t=1}^T [C_{t-1} - \sum_{i=1}^n c_{ti} y_i] \rho_t^h + \sum_{k=1}^K [D_k - \sum_{i=1}^n d_{ki} y_i] \lambda_k^h + \sum_{t=1}^{T-1} B_t \theta_t^h$$

$$+ \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n \{y_i u_{ij}^h + y_j v_{ij}^h + (1-y_i-y_j)w_{ij}^h\}$$

for  $h=1, \dots, H^0$  ;

(iii) 
$$\sum_{i \in S} y_i \leq 1 \quad \text{for each set } S \text{ of mutually exclusive projects ;}$$

(iv) 
$$y_i - y_j \leq 0 \quad \text{for each contingent project (i)-independent project (j) pair ;}$$

(v) 
$$y_i = 0 \text{ or } 1 \quad \text{for each } i .$$

(8.19)

The last  $n$  constraints of the restricted pure integer program in (8.19) simply call for the discreteness of the individual projects and cash-disbursement activities. The constraints in groups (iii) and (iv) express the mutual-exclusion and contingency relationships that exist among the candidate proposals. The restrictions in (iii)-(v) of the constraint set are thus simply the constraints involving only real projects and disbursement activities that appear in the complete model of the capital-budgeting problem (8.10).

Constraints of the type in (ii) are based upon extreme rays that have been enumerated up to the present step. This type of constraint was just discussed when (8.14) was presented. The constraints appearing in (i) of the restricted integer subproblem result from previously enumerated extreme points. Each of these extreme points was the optimal vector for a linear program of the type in (8.12) that had a finite solution value. Each such linear program was based on a  $y$ -vector that solved a previous restricted integer subproblem but that was not found optimal, at that stage, for the complete mixed-integer problem (8.10).

The qualification "at that stage" is needed because as more constraints are added to the restricted integer program, the maximum value of  $\Psi$  in that problem will decrease -- or at best stay constant. That is, the optimal value of  $\Psi$  in (8.19) is a nonincreasing function of the total number of extreme-ray and extreme-point constraints enumerated. Hence, we may find that while  $(\Psi^{M^0+H^0}, \bar{y})$

solves the  $M^0+H^0$ th restricted integer subproblem, the inequality

$$\Psi^{M^0+H^0} > \sum_{i=1}^n Q_i \bar{y}_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n P_{ij} \bar{x}_{ij} \quad \text{obtains so that } \bar{y} \text{ cannot be judged optimal}$$

for the complete problem (8.10). But as more constraints are added to (8.19) -- that is, as more extreme rays and extreme points are enumerated -- and the optimal value of  $\Psi$  decreases, we may reach a new restricted integer subproblem, say the  $M'+H'$ th, in which  $\bar{y}$  again emerges as the optimal solution vector, but now

with  $\Psi^{M'+H'} \leq \sum_{i=1}^n Q_i \bar{y}_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n P_{ij} \bar{x}_{ij}$ . This can clearly occur because the

expected utility generated by  $\bar{y}$  remains constant at  $\sum_{i=1}^n Q_i \bar{y}_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n P_{ij} \bar{x}_{ij}$ ,

while the optimal value of the restricted integer subproblem declines from  $\Psi^{M'+H^0}$  to  $\Psi^{M'+H'}$  as the firm recognizes more of the constraints it actually faces, that is, as the constraint set of (8.19) gets larger. In short, it may be found that  $\bar{y} = \{\bar{y}_i\}$  is, in fact, the optimal capital-investment program although it could not be judged optimal on the basis of the information available at the  $M^0+H^0$ th step.

In order to interpret these type (i) constraints, suppose a 0-1 subproblem has just yielded the solution vector  $\bar{y}$ . Suppose further that the resulting linear program in (8.12) has produced the solution vector  $\{\{\bar{\rho}_t\}, \{\bar{\lambda}_k\}, \{\bar{\theta}_t\}, \{\bar{u}_{ij}\}, \{\bar{v}_{ij}\}, \{\bar{w}_{ij}\}\}$  with a finite solution value that shows, however, that  $\bar{y}$  cannot now be judged optimal for the complete mixed-integer program (8.10). The new constraint generated is

$$\Psi \leq \sum_{i=1}^n Q_i y_i + \sum_{t=1}^T [C_{t-1} - \sum_{i=1}^n c_{ti} y_i] \bar{\rho}_t + \sum_{k=1}^K [D_k - \sum_{i=1}^n d_{ki} y_i] \bar{\lambda}_k + \sum_{t=1}^{T-1} B_t \bar{\theta}_t + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n [y_i \bar{u}_{ij} + y_j \bar{v}_{ij} + (1-y_i - y_j) \bar{w}_{ij}] . \quad (8.20)$$

This constraint may be rewritten as

$$\Psi \leq \sum_{t=1}^T C_{t-1} \bar{\rho}_t + \sum_{k=1}^K D_k \bar{\lambda}_k + \sum_{t=1}^{T-1} B_t \bar{\theta}_t + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n \bar{w}_{ij} + \sum_{i=1}^n \{Q_i - \sum_{t=1}^T c_{ti} \bar{\rho}_t - \sum_{k=1}^K d_{ki} \bar{\lambda}_k + \sum_{\substack{j \\ i < j}} (\bar{u}_{ij} - \bar{w}_{ij}) + \sum_{\substack{j \\ i > j}} (\bar{v}_{ji} - \bar{w}_{ji})\} y_i . \quad (8.21)$$

The constant term is the value of those resources the firm possesses over its investment horizon, independent of the particular investment program it chooses, and upon which the artificial projects, borrowing, and lending make demands or to which they make contributions. The total value of these resources is calculated by first multiplying each resource's quantity by the (accounting) value assigned it in the optimal solution to the dual linear program based on investment program  $\bar{y}$ . Then one sums these products. Each of the dual "values" used measures the particular resource's marginal contribution to the part of the investment program's expected utility that derives from interactions among

projects, that is, to  $\sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n P_{ij} x_{ij}$ , given that the firm pursues the program  $\bar{y}$ .

A resource's dual value in this case answers the question "What would be added to the investment program's expected utility due to changes in the expected utility of project interactions alone, if ceteris paribus an additional marginal unit of this resource could be obtained, assuming the firm is following the capital budget described by  $\bar{y}$ ?"

The resources that meet both conditions given above -- (i) that the firm owns them no matter what investment program it undertakes and (ii) that the artificial projects, borrowing, and lending make some demand upon or contribution to them -- include the budget dollars in each period, the firm's other limited resources, the firm's borrowing limits, and the set of unity upper bounds in the third constraint of each of the sets defining the  $x_{ij}$ -variables. The unity upper bound in the constraint  $y_i + y_j - x_{ij} \leq 1$ , in effect, represents a resource of the firm upon which individual real projects  $i$  and  $j$  make a demand and to which the artificial project  $ij$  contributes. An examination of a particular set of extreme-point constraints later will illustrate more clearly how these upper bounds act as scarce resources. The constant term in (8.21) is thus the accounting value (calculated as described above) of the firm's budget dollars,

other limited resources, borrowing limits, and unity upper bounds in  $y_i + y_j - x_{ij} \leq 1$  constraints, when these resources are priced on the assumption that the candidate investment program  $\bar{y}$  has been accepted.

The coefficient of the *i*th project's variable ( $y_i$ ) in (8.21) is its re-evaluated return. It is its revised return figure when all the firm's resources are priced on the assumption that  $\bar{y}$ , the investment program suggested by the most recently solved integer subproblem, has been accepted. The first part of the *i*th project's re-evaluated return is  $Q_i$ , the project's individual contribution to expected utility. It is the expected utility that results if project *i* and only project *i* is pursued. The second part of the revised return is the net opportunity cost to the firm of all corporate resources that project *i* requires, taking account of the fact that the *i*th proposal makes demands on some corporate resources and contributes to other resources. This net opportunity cost is computed using the dual prices obtained for the firm's resources in the capital-budgeting problem in artificial projects in (8.11).

Several features of this opportunity-cost calculation ought to be noted. First, project *i* gets penalized for the net demands it makes upon the scarce company dollars in the several periods,  $\sum_{t=1}^T c_{ti} \bar{\rho}_t$ , and the net demands it makes upon other limited company resources,  $\sum_{k=1}^K d_{ki} \bar{\lambda}_k$ . Since  $\bar{\rho}_t \geq 0$  and  $\bar{\lambda}_k \geq 0$  for all *t* and *k* respectively, if project *i* requires funds in the *t*th period ( $c_{ti} > 0$ ) or requires the *k*th type of resource ( $d_{ki} > 0$ ), it gets penalized (or at best gets zero charge) for its demand upon the particular financial or other type of resource. If the resource is a "free" one,  $\bar{\rho}_t = 0$  or  $\bar{\lambda}_k = 0$ , the project does not get charged for its use. But if  $\bar{\rho}_t > 0$  project *i* gets debited for its demand upon *t*th-period dollars while if  $\bar{\lambda}_k > 0$  it gets charged for its demand upon that *k*th resource of the firm. On the other hand, project *i*'s re-evaluated return will be increased if, in fact, it generates dollars for the firm in a period in which they are scarce ( $c_{ti} < 0$  when  $\bar{\rho}_t > 0$ ) or contributes some of the truly

scarce resource  $k$  ( $d_{ki} < 0$  when  $\bar{\lambda}_k > 0$ ). A system of penalties and bonuses thus takes account of the project's use of and contribution to the firm's financial and other resources, when  $\bar{y}$  is assumed to be the capital budget.

Second, remember that the extent to which an artificial project representing the interaction between project  $i$  and some other project  $j$  can be undertaken is constrained by the extent to which project  $i$  exists:  $x_{ij} \leq \bar{y}_i$  for  $i < j$  or  $x_{ji} \leq \bar{y}_i$  for  $i > j$ . The  $i$ th project thus itself constitutes a scarce corporate resource for the artificial-projects' capital-budgeting problem, (8.11). In determining the net opportunity cost of project  $i$  for revising its expected-utility return figure, project  $i$  receives credit for its contribution to this scarce company resource -- its own existence. If the interaction project is  $x_{ij}$ , that is, if  $i < j$ , project  $i$  receives a bonus of  $\bar{u}_{ij}$  while if the artificial project is  $x_{ji}$ ,  $i$  being greater than  $j$ , project  $i$  receives a bonus of  $\bar{v}_{ji}$ . Since  $u_{ij}$  and  $v_{ji}$  are both constrained to be nonnegative and since all  $\bar{u}_{ij}$ 's and  $\bar{v}_{ji}$ 's appear with positive signs in the coefficient of project  $i$  in (8.21), project  $i$  never gets penalized for contributing to its own existence.<sup>7</sup>

In contrast, as the re-evaluated return figure in (8.21) shows, project  $i$  is always penalized, or at best not charged, for its demand upon the firm's resource  $1 - \bar{y}_i - \bar{y}_j$  in the artificial-projects' capital-budgeting problem. If  $i < j$ ,  $\bar{w}_{ij}$  is subtracted from the  $i$ th project's individual contribution to expected utility while if  $i > j$ ,  $\bar{w}_{ji}$  is subtracted from the  $Q_i$ -figure in calculating project  $i$ 's revised return, and each  $w_{ij}$  ( $i < j$ ) and  $w_{ji}$  ( $i > j$ ) is constrained to be nonnegative.

Finally, note that the re-evaluation of project  $i$ 's return involves a summation of dual variables over all projects other than  $i$ . It entails adding  $\sum_{\substack{j \\ i < j}} (\bar{u}_{ij} - \bar{w}_{ij}) + \sum_{\substack{j \\ i > j}} (\bar{v}_{ji} - \bar{w}_{ji})$  to project  $i$ 's individual expected utility,  $Q_i$ , after

---

<sup>7</sup>Clearly, for each  $i, j$  pair one and only one of the pair  $\bar{u}_{ij}$  or  $\bar{v}_{ji}$  appears in the coefficient of  $y_i$ . The former appears if  $i < j$ , the latter if  $i > j$ .

the latter is adjusted for the project's demands on (or contributionsto) firm finances or other limited resources. This highlights the fact that the re-evaluated return of project  $i$  depends on whether or not project  $i$  is included in the presently considered budget and on what other projects are included in the candidate investment program. The revised expected-utility figure for project  $i$  is a function of the dual prices for the corporate resources that constrain the capital-budgeting decision in artificial projects. The revised figure for project  $i$  depends on the optimal values of the variables in (8.12). But these optimal values are clearly a function of which projects are included in and which projects are excluded from the proposed capital budget,  $\bar{y}$ . The examination below of a specific set of extreme-point constraints will help to emphasize this dependence of each project's revised valuation on the particular investment program the immediately preceding 0-1 subproblem has advised the firm to pursue.

In summary, then, the re-evaluated return of project  $i$  is a linear approximation to the change in the expected utility of the investment program that is effected if project  $i$ 's position in the present budget is altered, ceteris paribus. It equals the individual contribution of project  $i$  to the expected utility of the investment program adjusted for the value of the project's demands upon and contributions to the scarce resources of the firm, assuming the investment program  $\bar{y}$  has been accepted. If  $\bar{y}_i = 0$ , a positive re-evaluated return approximates the benefit, a negative one approximates the harm, that would result if project  $i$  were added to the budget with no other project's status being changed. If, on the other hand, project  $i$  is included in the proposed budget ( $\bar{y}_i = 1$ ), a negative revised return figure shows the approximate benefit that would result from dropping project  $i$  from the program, ceteris paribus, while a positive revised return figure shows the approximate loss in expected utility that would result if project  $i$  were dropped, again with all other projects' positions unchanged.

To see why the revised return figure for a project, as that figure appears in (8.21), provides a linear approximation to the change in expected utility resulting from a ceteris paribus change in the project's status, it is only necessary to appeal to basic duality theory. From that theory it follows that  $\bar{\rho}_t$  is the partial derivative of the expected utility resulting from interactions among the projects in the candidate program  $\bar{y}$  with respect to what is left for expenditure on artificial projects in the tth period, that is, with respect to

$$C_{t-1} - \sum_{i=1}^n c_{ti} \bar{y}_i . \text{ Similarly, } \bar{\lambda}_k \text{ is the partial derivative of } \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n P_{ij} \bar{x}_{ij} \text{ with}$$

respect to  $D_k - \sum_{i=1}^n d_{ki} \bar{y}_i$ ,  $\bar{u}_{ij}$  is the partial with respect to  $\bar{y}_i$  for  $i < j$ ,  $\bar{v}_{ji}$  is

the partial with respect to  $\bar{y}_i$  for  $i > j$ ,  $\bar{w}_{ij}$  is the partial with respect to  $1 - \bar{y}_i - \bar{y}_j$  for  $i < j$ , and  $\bar{w}_{ji}$  is the partial with respect to  $1 - \bar{y}_i - \bar{y}_j$  for  $i > j$ . Applying the chain rule for partial differentiation, the adjustment factor for the ith project,

$$- \sum_{t=1}^T c_{ti} \bar{\rho}_t - \sum_{k=1}^K d_{ki} \bar{\lambda}_k + \sum_{\substack{j \\ i < j}} (\bar{u}_{ij} - \bar{w}_{ij}) + \sum_{\substack{j \\ i > j}} (\bar{v}_{ji} - \bar{w}_{ji}) , \text{ be it presently accepted or}$$

rejected, is seen to be simply the partial derivative, with respect to  $\bar{y}_i$ , of the expected utility resulting from project interactions in the candidate budget.

Hence, the adjustment factor for the ith project exactly measures the marginal change in the expected utility of project interactions for a marginal, ceteris paribus change in the ith project's presence.

There is, however, no guarantee that the partial derivative precisely

measures the resultant change in  $\sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n P_{ij} \bar{x}_{ij}$  for a ceteris paribus change in

project size as large as unity. There is no guarantee that as some  $y_i$  goes from zero to one or from one to zero that one or more of the dual variables of (8.11) will not change in value. Therefore, one cannot state that the adjustment factor for project i exactly measures the gain or loss in project-interactions' expected



utility that would accrue if that project's position in the investment program were completely altered (an accepted project dropped, a rejected project added) and no other change made. The adjustment factors of individual projects are only linear approximations to the gains or losses in project-interactions' expected utility effected by such full-scale ceteris paribus changes in their positions. Hence, the re-evaluated return figure for a project,  $Q_i$  plus the adjustment factor based upon its net demands upon scarce corporate resources, is only a linear approximation to the actual change in expected utility that would result if that project's status were changed, leaving all other elements of  $\bar{y}$  unchanged.

One may note the similarity between the re-evaluation of expected-utility returns of the real projects in applying Benders' partitioning procedure to the capital-budgeting problem and the revision of divisional-output profit coefficients in the application of the Dantzig-Wolfe decomposition algorithm to production problems of large corporations.<sup>8</sup> In the former case, each time a new extreme-point constraint is generated, a new set of revised expected-utility figures is generated for the projects. These figures adjust the projects' individual returns on the basis of the projects' demands upon or contributions to corporate resources under the regime of the previously proposed capital-investment program. In the case of the large corporation's production problem, the per-unit profit figures for the divisions' products are revised at each iteration. The revised profit coefficient for a product adjusts the return each unit of that output brings to the division. The adjustment forces the division to take account of the value to the rest of the corporation of the scarce inputs the particular output uses, when the corporate resources are valued on the basis of the basic feasible corporate production schedule considered last.

---

<sup>8</sup> See Dantzig (1963, Chapter 23) and Baumol and Fabian (1964). In Balinski (1963), Balinski describes the mathematical relationship between the Benders and Dantzig-Wolfe procedures, when Benders' is applied to a pure linear programming problem. As he writes, "The [Benders'] procedure, ..., may be said to be 'dual' to the Dantzig-Wolfe 'decomposition algorithm.'" Balinski (1963, p. 1).

Using the idea of re-evaluated expected-utility returns for real projects (and disbursement activities), each extreme-point constraint of the type in (8.21) can be interpreted as follows. Given the candidate capital budget  $\bar{y}$  obtained as the solution vector for some restricted 0-1 subproblem but found, at that stage, not to be optimal for the complete capital-budgeting problem, (8.10), determine the "value" or "price" of each corporate resource in the resulting program in artificial projects.<sup>9</sup> The optimal value of expected utility for the firm's investment program then must not be greater than (i) the "value" of the corporation's cash throw-offs, other limited resources, borrowing limits, and unity upper bounds in  $\bar{y}_i + \bar{y}_j - x_{ij} \leq 1$  constraints, plus (ii) the re-evaluated returns of the real projects and disbursement activities, where the "prices" used to calculate the values in (i) and (ii) are the ones determined in the dual to the artificial-projects' capital-budgeting problem.

### 8.3.C. An Alternative Interpretation of the Extreme-Point Constraints<sup>10</sup>

An alternative interpretation of these extreme-point constraints sheds greater light on their economic meaning. To present this interpretation, let  $I = \{i | \bar{y}_i = 1\}$ . That is,  $I$  is the set of indices of all projects accepted in the proposed investment program  $\bar{y}$ . Then the constraint in (8.21) may be rewritten as

<sup>9</sup>Recall that this includes the "value" or "price" of each real project's existence:  $u_{ij}$  or  $v_{ji}$  for project  $i$  in defining  $x_{ij}$ .

<sup>10</sup>This alternative way of writing the extreme-point constraints is similar to -- and was suggested by -- an interpretation of the Benders constraints appearing in the application of the partitioning procedure to a plant-location problem in Balinski (1965, p. 290) and Balinski and Wolfe (1963, pp. 11-13).

$$\begin{aligned}
 \Psi \leq & \sum_{i=1}^n Q_i \bar{y}_i + \sum_{t=1}^T [C_{t-1} - \sum_{i=1}^n c_{ti} \bar{y}_i] \bar{\rho}_t + \sum_{k=1}^K [D_k - \sum_{i=1}^n d_{ki} \bar{y}_i] \bar{\lambda}_k + \sum_{t=1}^{T-1} B_t \bar{\theta}_t \\
 & + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n [\bar{y}_i \bar{u}_{ij} + \bar{y}_j \bar{v}_{ij} + (1 - \bar{y}_i - \bar{y}_j) \bar{w}_{ij}] \\
 (8.22) \quad & + \sum_{i \in I} [Q_i - \sum_{t=1}^T c_{ti} \bar{\rho}_t - \sum_{k=1}^K d_{ki} \bar{\lambda}_k + \sum_{\substack{j \\ i < j}} (\bar{u}_{ij} - \bar{w}_{ij}) + \sum_{\substack{j \\ i > j}} (\bar{v}_{ji} - \bar{w}_{ji})] [y_i - 1] \\
 & + \sum_{i \notin I} [Q_i - \sum_{t=1}^T c_{ti} \bar{\rho}_t - \sum_{k=1}^K d_{ki} \bar{\lambda}_k + \sum_{\substack{j \\ i < j}} (\bar{u}_{ij} - \bar{w}_{ij}) + \sum_{\substack{j \\ i > j}} (\bar{v}_{ji} - \bar{w}_{ji})] y_i .
 \end{aligned}$$

But the sum of the second, third, fourth, and fifth sums on the right-hand side of (8.22) constitutes the optimal value of the objective function of the linear program in (8.12) when  $y = \bar{y}$ . From duality theory, it follows that the sum of these terms thus equals the optimal value of the primal objective function in (8.11), namely  $\sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n P_{ij} \bar{x}_{ij}$ . The constraint in (8.22) is thus equivalent to

$$\begin{aligned}
 \Psi \leq & \sum_{i=1}^n Q_i \bar{y}_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n P_{ij} \bar{x}_{ij} \\
 (8.23) \quad & + \sum_{i \in I} [Q_i - \sum_{t=1}^T c_{ti} \bar{\rho}_t - \sum_{k=1}^K d_{ki} \bar{\lambda}_k + \sum_{\substack{j \\ i < j}} (\bar{u}_{ij} - \bar{w}_{ij}) + \sum_{\substack{j \\ i > j}} (\bar{v}_{ji} - \bar{w}_{ji})] [y_i - 1] \\
 & + \sum_{i \notin I} [Q_i - \sum_{t=1}^T c_{ti} \bar{\rho}_t - \sum_{k=1}^K d_{ki} \bar{\lambda}_k + \sum_{\substack{j \\ i < j}} (\bar{u}_{ij} - \bar{w}_{ij}) + \sum_{\substack{j \\ i > j}} (\bar{v}_{ji} - \bar{w}_{ji})] y_i .
 \end{aligned}$$

The sum of the first two terms in (8.23) is the expected utility of the present candidate investment program  $\bar{y}$ . This can be verified by re-examining the objective function of the full capital-budgeting problem in (8.10). The sum of the last two terms in (8.23) provides an upper bound on the change in expected utility that can be effected if projects are added to or removed from the present budget  $\bar{y}$  to yield a new budget  $y$ . Each extreme-point constraint of the type in

(8.21) or (8.23) states that the optimal capital budget's expected utility must be less than or equal to the value of the present candidate's expected utility plus an upper bound on the total change in expected utility that can be effected in moving from it.

This upper bound on the possible change in expected utility consists of two parts. The first part -- the third sum in (8.23) -- is associated with presently included projects while the second part -- the last sum in (8.23) -- is concerned with projects that are excluded from the present candidate investment plan. In each part of the upper bound on the change in expected utility, the coefficient of the ith project is the re-evaluated expected-utility figure that has just been discussed. Each project's coefficient -- be it a presently included or presently excluded proposal -- indicates the marginal change that would occur in total expected utility if that project's position and only that project's position were marginally changed in moving from  $\bar{y}$  to a new capital budget  $y$ . The upper-bound figure consisting of the last two terms in (8.23) is thus a linear approximation to what will most often be a nonlinear change in expected utility as one moves from investment program  $\bar{y}$  to investment program  $y$ .

The fact that the upper-bound figure only approximates (linearly) the true gain or loss in expected utility ought to be especially noted in the case in which more than one project's status is changed. The re-evaluation figure for project i is calculated on the assumption that all projects except the ith retain the status they had in the candidate investment program  $\bar{y}$ . Each of the re-evaluated returns used to compute the upper-bound figure is, in fact, a partial derivative and comprehends only the single-project change it was intended to measure. They cannot, when summed, take account of the possible interactions that may be gained or lost because more than one project's status is changed.

For example, suppose a project, A, presently in the budget possesses a negative re-evaluated return. This indicates that expected utility could be increased by dropping the project, ceteris paribus. Assume, furthermore, that a project, B, presently excluded from the budget simultaneously has a very high re-evaluated return, indicating that expected utility could be increased by a ceteris paribus addition of the project to the program. But simultaneously dropping A and adding B may not increase expected utility. Project A's negative revised return was calculated on the assumption that, for one thing, B was not in the budget while project B's high positive return was derived assuming, for one thing, that project A was in the budget. If projects A and B do, in fact, interact with highly beneficial results for the firm, simultaneously dropping A and adding B could be a detrimental action.

The myopic character of the upper bound's vision in this case results from the fact that the re-evaluated returns can only accurately gauge the impact of marginal single-project changes on expected utility. Other examples can easily be given that involve only accepted projects or only rejected projects. In the former case, consider two proposals that are presently accepted, each of which has a negative revised-return figure. If this is largely the result of strong negative interactions between the two of them -- which would be reflected, as shall be seen presently, in a high  $\bar{w}_{ij}$ -value -- dropping either one of them will be beneficial. Removing both of them from the budget, however, may sacrifice too much expected utility, either because of their individual returns or because of their positive interactions with other projects in the budget, and may, therefore, be a mistake. In the case of two presently rejected projects, suppose both of them have high positive revised-return figures. The myopic re-evaluation figures would seem to advise that both of them be added to the program. But, if the two projects have strong harmful interactions with one another, making this change could clearly be a mistake.

To summarize, then, each extreme-point constraint can be interpreted as imposing an upper bound on the expected utility of the optimal capital-investment program. It is bounded by the expected utility of the candidate budget upon which the constraint is based plus an upper bound on the possible change in expected utility that can be attained by moving from that candidate program, and this upper bound is a linear approximation to the change in expected utility that such a movement actually effects.

#### 8.4. Some Particular Aspects of the Solution Procedure

Having discussed the general extreme-point constraint (8.20), that appears in the restricted integer subproblem, now consider a particular set of extreme points upon which such constraints can be based. Specifically, consider the extreme points of the convex polyhedral set defined by the constraints in (8.12) in which only the  $u_{ij}$ ,  $v_{ij}$ , and  $w_{ij}$ -variables are possibly nonzero. That is, attention is focused upon nonnegative vectors  $\{\{\rho_t\}, \{\lambda_k\}, \{\theta_t\}, \{u_{ij}\}, \{v_{ij}\}, \{w_{ij}\}\}$  in which  $\rho_t = \lambda_k = \theta_t = 0$  for all  $t$  and  $k$ .

##### 8.4.A. A Justification for Examining These Extreme-Point Constraints

There are several reasons why these particular extreme points merit special attention. First, computationally, when the dual linear program in (8.12) has a finite solution, it always has a basic optimal solution with all but the  $u_{ij}$ ,  $v_{ij}$ , and  $w_{ij}$ -variables equal to zero. That is, if the minimization program in (8.12) possesses an optimum, there always exists a degenerate optimal solution vector with  $\{\rho_t\} = \{\lambda_k\} = \{\theta_t\} = 0$ , as will presently be demonstrated by construction.<sup>11</sup> Since these extreme-point solutions are so easily determined, as shall be seen, their existence suggests a convenient computational strategy. Namely, when a candidate budget  $\bar{y}$  admits a feasible solution vector

<sup>11</sup>See equation (8.28) and the discussion that follows it.

$\{x_{ij}\}, \{b_t\}, \{q_t\}$  to the maximization program in (8.11) but cannot be judged optimal for the full capital-budgeting model, (8.10), use the extreme points that have  $\{\rho_t\} = \{\lambda_k\} = \{\theta_t\} = 0$  to generate new extreme-point constraints. Then, only if the candidate budget appears again as the solution to a later restricted integer subproblem, and still cannot be judged to be the overall optimum at that time will we have to solve the full linear program in (8.12). This should considerably simplify the computational effort required in solving the full capital-budgeting problem.

A second justification for considering the particular set of extreme-point constraints in which  $\rho_t = 0$  for  $t=1, \dots, T$ ,  $\lambda_k = 0$  for  $k=1, \dots, K$ , and  $\theta_t = 0$  for  $t=1, \dots, T-1$ , rests with the nature of the capital-budgeting problem of a special type of firm. Specifically, if the corporation in question has neither borrowing nor lending opportunities -- that is, if it exists in a world of pure capital rationing -- and if, moreover, there are no cash-outflow or scarce-resource interactions among projects, the capital-budgeting problem assumes a form in which the only extreme points that exist, in fact, have  $\{\rho_t\} = \{\lambda_k\} = \{\theta_t\} = 0$ . To see this, note that in the absence of borrowing and lending opportunities and in the absence of interactions among projects with regard to constraining financial and physical resources, the model of capital budgeting under risk in (8.10) becomes

(8.24) {

$$\text{Maximize } \sum_{i=1}^n Q_i y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n P_{ij} x_{ij}$$

Subject to:

$$\sum_{i=1}^n c_{ti} y_i \leq C_{t-1} \quad \text{for } t=1, \dots, T .$$

$$\sum_{i=1}^n d_{ki} y_i \leq D_k \quad \text{for } k=1, \dots, K .$$

$$\sum_{i \in S} y_i \leq 1 \quad \text{for each set } S \text{ of mutually exclusive projects.}$$

$$y_i - y_j \leq 0 \quad \text{for each contingent project (i)-independent project (j) pair .}$$

$$\left. \begin{array}{l} -y_i + x_{ij} \leq 0 \\ -y_j + x_{ij} \leq 0 \\ y_i + y_j - x_{ij} \leq 1 \end{array} \right\} \text{for each } i, j \text{ pair with } i < j .$$

$$y_i = 0 \text{ or } 1 \quad \text{for each } i .$$

$$x_{ij} \geq 0 \quad \text{for each } i, j \text{ pair with } i < j .$$

Of course, in the absence of a lending alternative, the cash-disbursement activities numbered among the  $y_i$ -variables simply involve setting aside a fixed amount at the start of a period and disbursing that same amount at the end of the period. In addition, without the lending cash-carry-forward alternative, one might want to include a regular cash-carryover activity among the  $y_i$ -variables. Such an activity would simply allow a fixed sum to be carried forward from period  $t$ 's budget to period  $t+1$  where it could supplement the available funds.

The linear subproblems, when the capital-budgeting model is the mixed-integer program in (8.24), are given by (8.25) and (8.26), respectively.



$$(8.25) \left\{ \begin{array}{l} \text{Maximize } \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n P_{ij} x_{ij} \\ \text{Subject to: } \left. \begin{array}{l} x_{ij} \leq \bar{y}_i \\ x_{ij} \leq \bar{y}_j \\ -x_{ij} \leq 1 - \bar{y}_i - \bar{y}_j \end{array} \right\} \text{ for each } i, j \text{ pair with } i < j . \\ \\ x_{ij} \geq 0 \text{ for each } i, j \text{ pair with } i < j . \end{array} \right.$$

$$(8.26) \left\{ \begin{array}{l} \text{Minimize } \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n \{ \bar{y}_i u_{ij} + \bar{y}_j v_{ij} + (1 - \bar{y}_i - \bar{y}_j) w_{ij} \} \\ \text{Subject to: } \left. \begin{array}{l} u_{ij} + v_{ij} - w_{ij} \geq P_{ij} \\ u_{ij} \geq 0, v_{ij} \geq 0, w_{ij} \geq 0 \end{array} \right\} \text{ for each } i, j \text{ pair} \\ \text{with } i < j . \end{array} \right.$$

One very important difference between this special capital-budgeting model from which borrowing, lending, and resource-demand interactions between projects are absent and the more general model presented in (8.10) is the absence of any "extreme-ray problem" in the present model. Examining (8.25), it is clear that any  $\bar{y}$ -vector of zeros and ones admits a feasible  $\{x_{ij}\}$ -vector to (8.25). But any candidate budget that emerges as the solution to the restricted integer subproblem will necessarily be a vector of zeros and ones. Hence, any candidate budget results in a feasible maximization linear program in (8.25) and a finite solution value for the dual program in (8.26). The unique feasible  $\{x_{ij}\}$ -vector is determined by the constraints in (8.25) given the  $\bar{y}_i$ -values, and the objective function in (8.25) simply provides the value of that solution; there is no optimizing. Since the minimization program's solution value is thus always finite, solving the linear subproblems for this special case never leads to new extreme-ray constraints, but always to new extreme-point constraints, if the candidate budget is not judged optimal at the particular step.

In order to complete the description of the application of Benders' method to this special case of (8.10), note that the restricted integer subproblem takes the form in (8.27), where it is assumed that  $M^0$  extreme points of the convex polyhedral set defined by the constraints in (8.26) have been enumerated up to the present.

(8.27) {

$$\begin{aligned} & \text{Maximize } \Psi \\ & \text{Subject to: } \Psi \leq \sum_{i=1}^n Q_i y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n \{y_i u_{ij}^m + y_j v_{ij}^m + (1 - y_i - y_j) w_{ij}^m\} \\ & \hspace{15em} \text{for } m=1, \dots, M^0. \\ & \sum_{i=1}^n c_{ti} y_i \leq C_{t-1} \quad \text{for } t=1, \dots, T. \\ & \sum_{i=1}^n d_{ki} y_i \leq D_k \quad \text{for } k=1, \dots, K. \\ & \sum_{i \in S} y_i \leq 1 \quad \text{for each set } S \text{ of mutually exclusive projects.} \\ & y_i - y_j \leq 0 \quad \text{for each contingent project (i)-independent} \\ & \hspace{10em} \text{project (j) pair.} \\ & y_i = 0 \text{ or } 1 \quad \text{for each } i. \end{aligned}$$

In contrasting the solution procedure for the more general model with the procedure for the present special case, note several things. First, the procedure for the former model (8.10) contains extreme-ray constraints in the integer subproblem while such restraints are absent from the special case's integer subproblem (8.27). Second, for the more general model, the constraints on financial and other scarce resources appear in the linear subproblem while they appear in the integer subproblem of the special-case model, since in the latter these constraints involve only real projects. Finally, the only dual variables appearing in the model for pure capital rationing when there are no constraint interactions among projects are the  $u_{ij}$ ,  $v_{ij}$ , and  $w_{ij}$ -variables.

A last justification that can be offered for considering the particular set of extreme-point solutions of (8.12) in which each  $\rho_t$ ,  $\lambda_k$ , and  $\theta_t$  is zero rests with the interesting interpretation one can lend to the nonzero dual variables in these extreme points. Examining this particular set of extreme-point solutions and the constraints they generate will serve to illustrate and emphasize certain aspects of the previous discussion of the constraints of the type in (8.20).

#### 8.4.B. The Set of Special Extreme-Point Constraints

Suppose, then, that the preceding 0-1 subproblem of the type in (8.19) has just yielded the solution vector  $\bar{y}$ . In addition, assume the feasibility check described earlier leads to the conclusion that  $\bar{y}$  does admit a feasible  $\{\{x_{ij}\}, \{b_t\}, \{\ell_t\}\}$ -vector for the maximization linear program in (8.11). But comparing the optimal value of  $\Psi$  in the integer subproblem (8.19) that yielded  $\bar{y}$

with  $\sum_{i=1}^n Q_i \bar{y}_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n P_{ij} \bar{x}_{ij}$  reveals that  $\bar{y}$  cannot be judged optimal for the

entire capital-budgeting problem (8.10) at this time. An optimal solution vector for the dual linear program (8.12) is then given by:  $\bar{\rho}_t = 0$  for all  $t=1, \dots, T$ ,  $\bar{\lambda}_k = 0$  for all  $k=1, \dots, K$ ,  $\bar{\theta}_t = 0$  for all  $t=1, \dots, T-1$ , and case by case for  $i, j$  pairs, as follows:

$$(8.28) \left\{ \begin{array}{l} \text{(a) } \bar{y}_i = \bar{y}_j = 0: \text{ (i) } \bar{u}_{ij} = \max(0, P_{ij}), \bar{v}_{ij} = 0, \bar{w}_{ij} = 0 \\ \text{OR (ii) } \bar{u}_{ij} = 0, \bar{v}_{ij} = \max(0, P_{ij}), \bar{w}_{ij} = 0 \\ \text{(b) } \bar{y}_i = 1, \bar{y}_j = 0: \text{ (i) } \bar{u}_{ij} = 0, \bar{v}_{ij} = \max(0, P_{ij}), \bar{w}_{ij} = 0 \\ \text{OR (ii) } \bar{u}_{ij} = 0, \bar{v}_{ij} = \max(0, P_{ij}), \bar{w}_{ij} = \max(0, -P_{ij}) \\ \text{(c) } \bar{y}_i = 0, \bar{y}_j = 1: \text{ (i) } \bar{u}_{ij} = \max(0, P_{ij}), \bar{v}_{ij} = 0, \bar{w}_{ij} = 0 \\ \text{OR (ii) } \bar{u}_{ij} = \max(0, P_{ij}), \bar{v}_{ij} = 0, \bar{w}_{ij} = \max(0, -P_{ij}) \\ \text{(d) } \bar{y}_i = \bar{y}_j = 1: \text{ (i) } \bar{u}_{ij} = \max(0, P_{ij}), \bar{v}_{ij} = 0, \bar{w}_{ij} = \max(0, -P_{ij}) \\ \text{OR (ii) } \bar{u}_{ij} = 0, \bar{v}_{ij} = \max(0, P_{ij}), \bar{w}_{ij} = \max(0, -P_{ij}) \end{array} \right.$$

With  $\{\rho_t\} = \{\lambda_k\} = \{\theta_t\} = 0$ , the constraints in sets (ii), (iii), and (iv) of (8.12), specifically, constraints of the form

$\rho_t - (1+r_L)\rho_{t+1} \geq 0$ ,  $-\rho_t + (1+r_B)\rho_{t+1} + \theta_t \geq 0$ , and nonnegativity restrictions on each  $\rho_t$ ,  $\lambda_k$ ,  $\theta_t$ , respectively, are clearly satisfied. Moreover, with these variables set to zero, the objective function in (8.12) becomes

$$\text{Minimize } \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n \{ \bar{y}_i u_{ij} + \bar{y}_j v_{ij} + (1 - \bar{y}_i - \bar{y}_j) w_{ij} \},$$

and the remainder of the constraint set in (8.12) becomes

$$\left. \begin{array}{l} u_{ij} + v_{ij} - w_{ij} \geq P_{ij} \\ u_{ij} \geq 0, v_{ij} \geq 0, w_{ij} \geq 0 \end{array} \right\} \text{ for each } i, j \text{ pair with } i < j.$$

In short, the dual linear program in (8.12) reduces, as noted earlier, to the dual linear program of the model of pure capital rationing without interactions appearing in the constraints, that is, to the linear program in (8.26).

The solution in (8.28) clearly satisfies the nonnegativity restrictions in (8.26) since each  $u_{ij}$ ,  $v_{ij}$ , and  $w_{ij}$  is nonnegative. Turning to the constraints  $u_{ij} + v_{ij} - w_{ij} \geq P_{ij}$ , consider first case (a) in which  $\bar{y}_i = \bar{y}_j = 0$ . Using either alternative (i) or (ii), one obtains  $u_{ij} + v_{ij} - w_{ij} = \max(0, P_{ij})$  which is equal to  $P_{ij}$  if  $P_{ij} \geq 0$  and greater than  $P_{ij}$  if  $P_{ij} < 0$ . In case (b), with  $\bar{y}_i = 1$ ,  $\bar{y}_j = 0$ , alternative (i) leads to  $u_{ij} + v_{ij} - w_{ij} = \max(0, P_{ij})$  which is, again, greater than or equal to  $P_{ij}$  while alternative (ii) leads to  $u_{ij} + v_{ij} - w_{ij} = P_{ij}$ . Case (c) need not be examined separately since it is simply case (b) with the values of  $u_{ij}$  and  $v_{ij}$  reversed. Finally, in case (d), both alternatives, (i) and (ii), lead to  $u_{ij} + v_{ij} - w_{ij} = P_{ij}$ . Hence, for any  $i, j$  pair, the solutions in (8.28) combined with  $\{\rho_t\} = \{\lambda_k\} = \{\theta_t\} = 0$  satisfy the constraints of (8.12).

It only remains to show that the proposed solutions lead to an objective-function value in (8.12) equal to  $\sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n P_{ij} \bar{x}_{ij}$ , the optimal solution value of

(8.11). From the definition of the  $x_{ij}$ -variables -- that is, from the constraints in (iv)-(vii) in (8.11) -- it is known that  $\sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n P_{ij} \bar{x}_{ij}$  will exactly equal

the sum of the  $P_{ij}$ -values for all  $i, j$  pairs with  $i < j$  and  $\bar{y}_i = \bar{y}_j = 1$ . But the contribution of any  $i, j$  pair with  $i < j$  to the objective function of the linear program in (8.26) -- the one to which (8.12) reduces when each  $\rho_t, \lambda_k, \theta_t$  is set equal to zero -- is  $\bar{y}_i u_{ij} + \bar{y}_j v_{ij} + (1 - \bar{y}_i - \bar{y}_j) w_{ij}$ . Examining this contribution case by case in (8.28), one finds it is equal to zero in all but case (d). In this last case, when  $\bar{y}_i = \bar{y}_j = 1$ , the objective-function contribution is equal to  $u_{ij} + v_{ij} - w_{ij}$  which equals  $P_{ij}$  using either (i) or (ii) in (8.28)(d). Hence, only when  $\bar{y}_i = \bar{y}_j = 1$  does the set of dual variables corresponding to the  $i, j$  pair contribute a nonzero amount to the dual objective function in (8.26), and then it contributes exactly  $P_{ij}$ . Thus, combining  $\rho_t = 0, \lambda_k = 0,$  and  $\theta_t = 0$  for all  $t$  and  $k$  with the solution given in (8.28), the value of the dual objective function in (8.12) does indeed equal the sum of the  $P_{ij}$ -values for all  $i, j$  pairs with  $i < j$  and  $\bar{y}_i = \bar{y}_j = 1$ ; that is, it equals  $\sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n P_{ij} \bar{x}_{ij}$ . The solution

given by (8.28) and  $\{\rho_t\} = \{\lambda_k\} = \{\theta_t\} = 0$  is feasible for (8.12) and equates the primal (8.11) and dual (8.12) objective functions: it is an optimal solution for the linear program in (8.12).

Note that for each  $i, j$  pair with  $i < j$ , there are two possible solutions in (8.28) for the triplet  $\bar{u}_{ij}, \bar{v}_{ij},$  and  $\bar{w}_{ij}$  for the given candidate program  $\bar{y}$ . Hence, there exists a number of optimal extreme-point solutions to the dual linear program in (8.12) that have all  $\rho_t, \lambda_k,$  and  $\theta_t$ -variables equal to zero. Depending on computational considerations, the first time a particular investment

program  $\bar{y}$  appears as the candidate budget, that is, as the solution to a restricted integer subproblem, one may decide to add only a single constraint to the restricted zero-one subproblem, several constraints, or all the constraints that can be obtained from this multiplicity of optimal extreme-point solutions with  $\{\rho_t\} = \{\lambda_k\} = \{\theta_t\} = 0$ . To reach the global optimum of the full capital-budgeting problem in (8.10), it may, however, eventually be necessary to enumerate and include all the extreme-point constraints that can be derived from solutions to (8.12) with the given  $\bar{y}$  as the candidate budget. This full set of extreme points based on  $\bar{y}$  includes: (i) all the extreme-point solutions to (8.12) with  $\{\rho_t\} = \{\lambda_k\} = \{\theta_t\} = 0$  and  $u_{ij}$ ,  $v_{ij}$ , and  $w_{ij}$ -values obtained from (8.28), and (ii) all the extreme-point solutions to (8.12) in which not all  $\rho_t$ ,  $\lambda_k$ ,  $\theta_t$  are zero and which can only be obtained by solving the full linear program in (8.12). Exactly how many constraints derived from extreme-point solutions to (8.12), when the latter is based upon a particular  $\bar{y}$ , will eventually be required depends on the particular problem at hand.

With regard to the number of extreme-point constraints available when the linear program (8.12) is solved with  $\{\rho_t\} = \{\lambda_k\} = \{\theta_t\} = 0$  for a given  $\bar{y}$ , the solutions in (8.28) make it appear at first that there are  $\frac{n(n-1)}{2}$  of them available. (The number  $\frac{n(n-1)}{2}$  derives from the fact that there appear to be two alternatives for each  $i, j$  pair with  $i < j$ ; and there are  $\frac{n(n-1)}{2}$  such  $i, j$  pairs.) But this is not necessarily the case. First, if both projects  $i$  and  $j$  are rejected or if both are accepted, equation sets (8.28) (a) and (d) show, respectively, that the two solutions, (i) and (ii), given for the  $i, j$ -pair variables are identical if the projects have negative interactions with one another. This follows because with  $P_{ij} < 0$ ,  $\max(0, P_{ij}) = 0$ . At the same time, equation sets (8.28)(b) and (c) show that if one and only one member of the  $i, j$ -pair is included in the candidate budget and if the projects

have a beneficial interaction with each other, the two solutions given for each of these cases are identical. This occurs because  $P_{ij} > 0$  implies that  $\max(0, -P_{ij}) = 0$ . In short, when solving (8.12) there will generally be more than one extreme-point solution but fewer than  $2^{\frac{n(n-1)}{2}}$  extreme-point solutions that have  $\{\rho_t\} = \{\lambda_k\} = \{\theta_t\} = 0$ .

The ease with which such extreme-point solutions can be obtained, just by examining the status of each  $i, j$  pair and without any use of the simplex method, should be clear. It seems that using the computing strategy suggested earlier, with these extreme points and their corresponding constraints at its center, would be a computationally efficient way to proceed within the framework of Benders' algorithm. If such a strategy were used, the constraints added to the restricted integer subproblem each time a candidate budget  $\bar{y}$  made its first appearance would take the form

$$(8.29) \quad \Psi \leq \sum_{i=1}^n Q_i y_i + \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n \{y_i \bar{u}_{ij} + y_j \bar{v}_{ij} + (1-y_i-y_j) \bar{w}_{ij}\}.$$

The constraint in (8.29), which is also exactly the typical extreme-point constraint of the pure-integer problem in (8.27), may be rewritten as

$$(8.30) \quad \Psi \leq \sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n \bar{w}_{ij} + \sum_{i=1}^n \{Q_i + \sum_{\substack{j \\ i < j}} (\bar{u}_{ij} - \bar{w}_{ij}) + \sum_{\substack{j \\ i > j}} (\bar{v}_{ji} - \bar{w}_{ji})\} y_i.$$

The constant term in this extreme-point constraint is the value to the firm of the only truly scarce (not "free") resources (i) that it owns no matter what investment program it chooses and (ii) that the artificial projects, borrowing, or lending require or help create. Since  $\bar{\rho}_t = 0$  for all  $t$ ,  $\bar{\lambda}_k = 0$  for all  $k$ , and  $\bar{\theta}_t = 0$  for all  $t$ , the financial budgets, other corporate

resources, and borrowing limits are actually "free" not "economic" goods -- they each have a zero dual price. The only resources meeting conditions (i) and (ii) above are the unity upper bounds in the third constraint of each set defining an  $x_{ij}$ -variable in (8.10):  $y_i + y_j - x_{ij} \leq 1$ . The price of the  $ij$ th upper-bound scarce "resource" is  $\bar{w}_{ij}$ .

From duality theory and an examination of the maximization linear program in (8.11), it follows that  $\bar{w}_{ij}$  measures the marginal increase in the expected utility of project interactions that would result if the  $ij$ th unity upper bound could be increased marginally, given  $\bar{y}$  as the candidate investment program. Since  $w_{ij}$  is constrained to be nonnegative, either the unity upper bound is positively valued for benefits an increase in it could effect or it receives a zero price. But an increase in the  $ij$ th unity upper bound would allow a decrease in  $x_{ij}$ , the artificial project capturing expected-utility interactions between projects  $i$  and  $j$ . One suspects, then, some relationship between  $w_{ij}$  and undesirable interaction effects since  $w_{ij}$  awards a positive price to a "resource" that would allow less of the artificial project to be accepted. This is precisely what the solutions given in (8.28) indicate for  $w_{ij}$  is always either equal to zero (always so in case (a) where both real projects are rejected) or equal to the magnitude of the negative (detrimental) interaction effect of the two projects. That is,  $\bar{w}_{ij}$  is always equal to zero or the  $\max(0, -P_{ij})$ , and the latter is only positive when  $P_{ij} < 0$ .

The sum  $\sum_{i=1}^n \sum_{\substack{j=2 \\ i < j}}^n \bar{w}_{ij}$  thus measures the value to the firm of the unity

upper bounds in the  $y_i + y_j - x_{ij} \leq 1$  constraints. The sense in which these upper bounds are scarce resources to the firm is hopefully clearer now. If the  $ij$ th unity upper bound could be increased, the firm would gain nothing if projects  $i$  and  $j$  interacted to the firm's benefit. But the firm would gain expected utility



from such an increase if the projects' interaction actually reduced expected utility because the increase in the upper bound would then let the corporation decrease the extent to which the harmful project  $x_{ij}$  was accepted.

Turning to the re-evaluated return figure for the  $i$ th project, its coefficient in (8.30), the first part of the revised return is project  $i$ 's individual contribution to expected utility,  $Q_i$ . The adjustment of the individual return to take account of the role project  $i$  plays in the expected utility of project interactions equals  $\sum_{\substack{j \\ i < j}} (\bar{u}_{ij} - \bar{w}_{ij}) + \sum_{\substack{j \\ i > j}} (\bar{v}_{ji} - \bar{w}_{ji})$ . Each  $i, j$  pair thus contributes either  $\bar{u}_{ij} - \bar{w}_{ij}$  or  $\bar{v}_{ji} - \bar{w}_{ji}$  (but not both) to the adjustment factor for project  $i$ . But the values assumed by  $\bar{u}_{ij}$  and  $\bar{w}_{ij}$  or  $\bar{v}_{ji}$  and  $\bar{w}_{ji}$  are, as the solutions in (8.28) make clear, dependent upon the positions of projects  $i$  and  $j$  in the candidate budget. Hence, each  $i, j$  pair's contribution to the re-evaluation of project  $i$  depends on whether or not project  $i$  or project  $j$  or both were in the candidate investment program  $\bar{y}$ . Since the total adjustment of the  $i$ th project's individual return equals the sum of the pair-by-pair adjustment factors, it is clear that the re-evaluated return of project  $i$  is a function of the exact composition of the proposed capital budget  $\bar{y}$ .

As in the general case, the adjustment factor for project  $i$  equals the partial derivative of the expected utility of project interactions with respect to  $\bar{y}_i$ . It measures the marginal change in the expected utility of project interactions for a ceteris paribus marginal change in the  $i$ th project's presence. Since the re-evaluation for any particular project  $i$  is, in fact, the sum of adjustments made on a project-pair basis, it is best to consider the nature of the re-evaluation process by examining how the individual returns of two projects ( $i$  and  $j$ , assume  $i < j$ ) are modified when they occupy the different possible positions in the present program,  $\bar{y}$ .

First, suppose neither  $i$  nor  $j$  is included at present. Then the adjustment factors are

$$(8.31) \quad \left\{ \begin{array}{l} \text{(i)} \quad \bar{u}_{ij} - \bar{w}_{ij} = \max(0, P_{ij}) \quad \underline{\text{and}} \quad \bar{v}_{ij} - \bar{w}_{ij} = 0 \\ \underline{\text{or}} \\ \text{(ii)} \quad \bar{u}_{ij} - \bar{w}_{ij} = 0 \quad \underline{\text{and}} \quad \bar{v}_{ij} - \bar{w}_{ij} = \max(0, P_{ij}). \end{array} \right.$$

In case (i), project  $i$  receives credit for the entire potential interaction if it is beneficial while it receives no blame for any potential interaction that may be harmful. Project  $j$ 's return is, at the same time, unmodified no matter whether the two projects interact to the good or ill of the investment program. In the set of adjustment factors based upon the alternative solution, (8.28)(a)(ii), the positions of projects  $i$  and  $j$  are simply reversed. It is important to note that no blame is assessed in either case to either of the projects for any harmful potential interaction between them.

If project  $i$  is accepted ( $\bar{y}_i = 1$ ) but  $j$  is rejected ( $\bar{y}_j = 0$ ) in the candidate program  $\bar{y}$ , the alternative sets of figures to use in calculating the total re-evaluation for the projects are

$$(8.32) \quad \left\{ \begin{array}{l} \text{(i)} \quad \bar{u}_{ij} - \bar{w}_{ij} = 0 \quad \underline{\text{and}} \quad \bar{v}_{ij} - \bar{w}_{ij} = \max(0, P_{ij}) \\ \underline{\text{or}} \\ \text{(ii)} \quad \bar{u}_{ij} - \bar{w}_{ij} = \min(0, P_{ij}) \quad \underline{\text{and}} \quad \bar{v}_{ij} - \bar{w}_{ij} = P_{ij}. \end{array} \right.$$

If the first set of adjustment factors is used, project  $i$  receives neither credit nor blame for any potential interaction it has with project  $j$ . The latter, on the other hand, receives credit -- its individual expected-utility figure is augmented by  $P_{ij}$  -- if the potential interaction is beneficial while it is not debited for any negative interaction,  $P_{ij} < 0$ . If the first solution in (8.28)(b)

is used when  $\bar{y}_i = 1$  and  $\bar{y}_j = 0$ , neither project's expected-utility return is ever decreased because of a detrimental interaction between them.

The story is, however, a different one if the extreme-point constraint generated by  $\bar{y}$  uses the second possible solution in (8.28)(b), which leads to set (ii) of the re-evaluation figures in (8.32). In this case, project  $i$  -- the included project -- never receives any credit for a potential interaction it has with  $j$  that is beneficial to the firm. It is penalized, however, by the full amount of the interaction effect,  $P_{ij}$ , if the potential interaction is harmful. This is clear because  $\min(0, P_{ij})$  will be nonzero if and only if  $P_{ij}$  is negative. The rejected project,  $j$ , in contrast, is assessed the full credit if  $P_{ij} > 0$  and the full blame if  $P_{ij} < 0$  for whatever potential interaction effect it and project  $i$  may have on the program's expected utility.

Consider more closely the debiting, in (8.32)(ii), of each project's return when they interact negatively. Decreasing the individual expected-utility figure of the presently rejected project is easily explained. If the presently rejected project  $j$  were to be accepted, ceteris paribus, it would interact with the  $i$ th project. But if the two projects interact to the detriment of the investment program, the expected utility of the program would decrease by  $P_{ij}$ . Hence, it makes sense to debit the individual expected-utility return of the  $j$ th project so as to indicate that a ceteris paribus change in its position from what it is in  $\bar{y}$  would not only increase expected utility by  $Q_i$  but also simultaneously decrease it by  $P_{ij}$ .

The reason for deducting the absolute value of  $P_{ij}$  from the individual return of the  $i$ th project rests on a different ground. Namely, since the second set of adjustment figures in (8.32) is based on the second solution in (8.28)(b), when this set is used, the scarce resource constituted by the unity upper bound in  $y_i + y_j - x_{ij} \leq 1$  is receiving a positive price if the two projects' interaction





accepted and the other rejected, (8.32)(i) and (ii) and (8.33)(i) and (ii), respectively, are, on the other hand, equal when the interaction between the pair results in an increase in expected utility.

The set of re-evaluation figures presented in (8.31)-(8.34) contains a peculiar mixture of optimism and pessimism with regard to the merit of individual projects. As the discussion of the more general extreme-point constraint (8.20) showed, the upper-bound figure each such constraint provides for the change in expected utility effected in moving from  $\bar{y}$  to  $y$  is but a linear approximation to the true change that takes place. When extreme-point solutions with  $\{\rho_t\} = \{\lambda_k\} = \{\theta_t\} = 0$  are used to generate such constraints, the adjustment figures that result in this linear approximation sometimes overvalue, sometimes undervalue individual projects.

An example of the optimism of these figures occurs when both projects  $i$  and  $j$  are presently rejected. Neither project receives any blame for any detrimental potential interaction that would occur if both of them were accepted. Although this is in accord with the marginal-analysis character of the re-evaluation figures, since the harm is not engendered unless both projects are accepted, it seems to overvalue each of the projects. This optimism also extends to the case of positive potential interactions between the two presently rejected projects. The project that receives the positive adjustment when  $P_{ij} > 0$ , receives credit for the full value of the interaction -- that is,  $P_{ij}$  is added to its individual return. But this positive interaction could not be effected unless both projects were accepted. Hence, this overstates the "value" of the project receiving the credit and understates the value of the other project, without whose joint presence the benefit of the interaction could not be obtained.

Such overvaluation and undervaluation also occurs in the case where one project is presently included in the budget and the other is not. If the first

set of adjustment figures is used -- (i) in (8.32) or (8.33) -- optimistically, neither project is debited for any detrimental interaction effects that might occur if the rejected project's status were changed. In addition, the presently rejected project gets full credit for any positive interaction that may be effected by introducing it into the program, and the presently accepted project gets no credit for it. The latter follows the dictates of marginal analysis for introducing the rejected project, ceteris paribus, would increase the expected utility of project interactions. But this allocation of the potential benefit undervalues the presently accepted project and overvalues the presently rejected one.

If the second set of adjustment factors is used when one project is in and the other out of the candidate program -- (ii) in (8.32) and (8.33) -- a similar but more pessimistic overvaluation and undervaluation takes place. Once again, the rejected project receives full credit for any  $P_{ij} > 0$  while the accepted project receives no credit for this joint benefit. On the other hand, if the interaction effect is negative,  $P_{ij} < 0$ , both projects are penalized in full for their harmful interaction. It would seem that both of them are undervalued in this last case since neither is fully responsible for their joint detrimental effect.

Finally, in the case where both projects i and j have been accepted in the most recently proposed budget, optimism and pessimism again appear in their re-evaluation. No matter whether the set of adjustment factors used is (i) or (ii) in (8.34), each project gets penalized by the full amount of any negative interaction effect they may jointly have. This clearly agrees with marginal analysis since removing either one of them, ceteris paribus, would increase the expected utility due to project interactions by the full value of  $P_{ij}$ . Yet such an adjustment factor seems to understate the "value" or "merit" of both projects since neither one is fully responsible for the harm they jointly cause. If,

however, the two projects interact to the benefit of the firm, one of them gets credited with the full value of the positive interaction effect while the other has no addition made to its individual expected-utility figure. Which project receives the entire credit depends on whether one bases the new constraint on (i) or (ii) in (8.34). In either case, the merit of the project credited with the interaction is overstated while the merit of the other is understated.

### 8.5. A Summary of the Iterative Decision Procedure

Despite this optimism and pessimism, this undervaluation and overvaluation of individual projects, and despite the fact that the upper-bound figures in the extreme-point constraints as the one in (8.20)-- and the special case (8.30) -- are but linear approximations to true changes in expected utility, the iterative procedure described in this chapter locates the globally optimal investment program. By applying Benders' partitioning procedure to the capital-budgeting problem of the firm facing the risks and the capital market described earlier, one can find the best allocation of that firm's funds. Let us summarize the iterative decision procedure that locates this optimal set of projects.

Begin by selecting any investment program, any subset of the candidate proposals; call it  $\bar{y}$ . Apply the feasibility check described and thereby determine whether or not there exists a feasible pattern of borrowing and lending and whether or not there exists enough of the resources  $D_k$  ( $k=1, \dots, K$ ) to enable the firm to pursue the capital-investment program proposed. If resources do not suffice or if there does not exist such a schedule of borrowing and lending, that is, if there does not exist a feasible vector  $\{\{x_{ij}\}, \{b_t\}, \{l_t\}\}$  to the program in artificial projects, borrowing, and lending, when  $\bar{y}$  is the candidate investment program, a constraint is generated for the integer subproblem that will suggest future candidate capital budgets. This constraint serves to exclude this particular combination of proposals,  $\bar{y}$ , from further consideration.



If, on the other hand,  $\bar{y}$  is feasible for the entire capital-budgeting problem, by solving the dual linear program in (8.12), with  $\bar{y}_i = 1$  or  $\bar{y}_i = 0$  as indicated by  $\bar{y}$ , one or more sets of re-evaluation figures are derived. Each set contains one figure for each real project. These adjustment figures are used to modify the independent contribution each project makes to expected utility,  $Q_i$ . The resulting re-evaluated return figure for the  $i$ th project shows the marginal change in the investment program's total expected utility that would result if the project's status were marginally changed from what it is in the candidate program. Each set of revised return figures is then used to derive a constraint involving only real-project variables. It states that the optimal value of total expected utility cannot exceed the total expected utility of the candidate program,  $\bar{y}$ , plus an upper bound on the possible change in expected utility that could be effected by moving from the candidate budget to a new program.

As a result of this first candidate investment program, there now exists a pure integer, actually 0-1, subproblem in the real projects alone. The integer program contains one constraint if  $\bar{y}$  was found to be infeasible for the overall capital-budgeting problem. It consists of one or more constraints if, on the other hand, it was found that  $\bar{y}$  could be financed by the firm and undertaken given the firm's other scarce resources. That is, this first restricted integer subproblem contains either one extreme-ray constraint or one or more extreme-point constraints, depending on whether  $\bar{y}$  was, in fact, infeasible or feasible for the capital-budgeting problem in (8.10).

At each step after this initial one a new candidate budget is not simply chosen at random but rather is located by solving a pure-integer programming problem that contains constraints of the type just described. Some of the constraints serve to remove previously proposed capital budgets as possibilities. The other members of the restricted 0-1 problems' constraint sets provide information about project interrelationships in the form of upper-bound constraints

on the optimal investment program's expected utility. The constraints in the integer subproblem are based on the solutions of the duals to previously generated linear programs in artificial projects, borrowing, and lending. Each of these linear programs has been generated in the same way as the initial one described above. Specifically, the  $y_i$ -variables corresponding to projects included in the most recently found candidate capital budget were each set equal to one; all other  $y_i$ -variables were set equal to zero.

When the optimal investment program for the restricted 0-1 subproblem is found, taking into account only the limited information about the feasibility of candidate budgets and only the limited information about project interactions contained in that subproblem, a new candidate capital budget is at hand. If it is infeasible for the complete capital-budgeting problem, a new extreme-ray constraint based on the solution to (8.12) is generated. It is added to the constraint set of the integer subproblem which is then solved once again.

If the candidate is feasible, using the objective function of the artificial-projects problem, (8.11), the relevant information is obtained concerning project interactions in the candidate program. This information enables us to decide whether or not this most recently proposed set of projects is optimal for the entire capital-budgeting problem. If the candidate program is not optimal, the dual to its artificial-projects subproblem, namely (8.12), yields one or more new extreme-point constraints in real projects alone. They are added to the 0-1 problem and the whole procedure is repeated.

But if the candidate budget is found to be optimal for the entire capital-budgeting problem, our task is completed. The optimal capital-investment program has been found.

## CHAPTER 9

### SUMMARY AND DIRECTIONS FOR FURTHER WORK

#### 9.1. Summary

The goal of this study was to construct a model of capital budgeting under risk for a firm operating in an imperfect capital market and to present a procedure for solving the resulting problem. This objective has been achieved. A model of the problem of allocating fixed budget dollars in several time periods among competing investment proposals has been formulated, and a solution procedure for that model has been presented for a firm facing the imperfect capital market and the risks described earlier. Specifically, subject to an absolute borrowing limit in each period, the firm could borrow and lend funds at constant rates of interest with the borrowing rate exceeding the lending rate. Insofar as the nondeterministic nature of its environment was concerned, the gross returns from potential investment projects were stochastic. Although this was the only uncertain feature of its opportunity set, as was indicated earlier, this is, in fact, a large part of the risk a firm faces when it chooses a capital budget.

A brief review of the development of the study may be helpful in indicating exactly where we have been and in providing some directions for where one might want to go from here. Since the capital-budgeting decision is a problem of decisionmaking over time in a risk environment, the study began with a general discussion of such decisionmaking situations. Attitudes toward risk in a multi-period context, in particular, risk aversion and decreasing risk aversion, were characterized in terms of properties of the decisionmaker's multiperiod utility

function. The new concept of risk balancing over time was introduced as an important member of the set of sufficient conditions for multiperiod decreasing risk aversion.

Temporarily assuming away the risk aspects of the environment, we considered the goal of the investment program of the firm. It was argued that the appropriate objective was maximization of a nonlinear utility function. The utility function was to be management's perception of the owners' utility as a function of the latter's consumption possibilities over the T periods of the firm's planning horizon. Since the true increases that the investment program generates in stockholders' consumption possibilities are not measurable, the amounts available for withdrawal from the firm in the several periods were taken as proxy measures of the former. Previously suggested approaches for investment-project selection, including previous programming approaches, were then shown to fall short of optimizing this objective function. A nonlinear programming approach with the nonlinear utility function as the objective function was seen to be the method to use in determining the firm's optimal capital budget.

The new nonlinear programming model that seemed called for in light of this examination of previous suggestions was presented next. A brief review of earlier surveys' findings on the way capital budgets are, in fact, drawn up in corporations served to highlight the normative, even utopian, nature of the programming model presented. The model is closely akin to previous programming approaches to the capital-budgeting decision. In some respects it is even identical with these earlier constructs. But it does have several distinctive features that are worth recalling at this time. First, the model places a nonlinear utility function as the maximand. Second, it is founded on a particular period-analytic view of the firm's cash-flow operations. Third, special attention is paid to the interactions among projects' cash flows and to the ways in which these interactions can be incorporated into the model. A specific

artificial-project approach to cash-flow interactions is adopted after consideration of the available alternatives. Fourth, the precise capital-market opportunities and disbursement activities present set the model apart from its predecessors.

After presenting the nonlinear programming model of capital budgeting under certainty, the question of capital budgeting under risk was re-opened. The specific risk environment assumed to exist in the present model was described, and previous suggestions for solving the problem of capital budgeting in this environment were evaluated. The finite-horizon approach, the risk-discount method, sensitivity analysis, chance-constrained programming, and the mean-variance approach were all found to be less than satisfactory as methods for leading the firm to an optimum in this environment.

A new programming approach to the problem of capital budgeting in the particular risk and imperfect-capital-market setting was presented. To begin with, the expected-utility hypothesis was taken as the basis for rational behavior under risk. A set of axioms that a multiperiod utility function serving as the capital-budgeting criterion function ought to satisfy was presented. These desiderata included positive marginal utility, boundedness, risk aversion, and decreasing risk aversion. It was also required that the utility function obviate the need to specify fully the joint frequency function of the withdrawals from the firm. A cubic utility function was shown to be the lowest-degree polynomial utility function satisfying these axioms, and it was also shown that a cubic that included a particular type of interaction between different periods' consumption incomes could also satisfy the axioms.

The discussion proceeded, for reasons of expositional simplicity, in terms of the cubic without interperiod interactions among consumption incomes. A decisionmaker applying the expected-utility theorem to a cubic utility function satisfying the axioms was shown to prefer ceteris paribus increases in expected

return and in positive (or rightward) skewness and ceteris paribus decreases in variance. It was demonstrated how the capital-budgeting objective function for this risk environment could be written in terms of the individual investment proposals. The artificial-project approach that was used in the certainty model was again used to capture cash-flow interactions, now stochastic ones, among projects.

The assumptions about the risk environment, intentionally, ensured that the constraint set faced by the firm under risk was the same as under certainty. The objective function that resulted from the axiomatic investigation was therefore combined with the constraint set of the certainty model to yield a programming model of capital budgeting in the risk and imperfect-capital-market environment described. The resulting model was a linear mixed-integer program.

Finally, a solution procedure based upon Benders' partitioning algorithm for mixed-variables problems was presented for solving the resulting capital-budgeting model. The algorithm's progress toward a solution was lent an economic interpretation, involving continual re-evaluation of individual projects' returns to take account of project interactions.

## 9.2. Directions for Further Work

There are two important directions in which further work might well proceed from here. The first involves theoretical extensions of what has been done in this study while the second direction entails the actual operational use of the model and procedure presented. Clearly, if the work done in this study is to be of help to capital-budgeting practitioners, the latter path of extension is not just desirable, it is imperative.

Considering the theoretical extensions first, there are several subject areas that the present study indicates merit further attention. First, there is the construction of a general-equilibrium model encompassing the simultaneous

solution of all individual and institutional investors' portfolio-selection problems and all corporate capital-budgeting problems. As noted several times in the course of the study, this is the truly correct setting in which to consider a firm's capital-budgeting decision. The work of J. Lintner<sup>1</sup> constitutes an auspicious start in the construction of such a model, but much remains to be done, particularly in the direction of relaxing several of his restrictive assumptions. Second, one would want to consider the problem of capital budgeting in a risk environment as a problem of sequential decisionmaking. Account would be taken not only of presently available or presently foreseen projects but also of those investment opportunities that may arise randomly through time. Further work along these lines might consider attempting to integrate the work presented here with work of the type done by G. M. Kaufman.<sup>2</sup>

Further generalization of the model developed would be made possible, as the assumptions imposed for reasons of tractability have indicated, by a number of developments in integer programming and stochastic programming. With respect to the former, methods for solving problems containing both integer-valued and continuous-valued variables with nonlinearity in both sets of variables and with products of integer and continuous variables would certainly be of great help. For example, one could then analyze a model that contained both discrete investment proposals and continuous disbursement options. In stochastic programming, at several points in the discussion we have noted the desirability of having a satisfactory way for coping with stochastic elements in the constraint matrix. Such a development would enable relaxation of the assumptions that gross cash outlays are nonstochastic and that none of the stochastic net returns from the investment program are reinvested in the firm. Since this list of desirable extensions is asking for quite a bit, one might as well get slightly more.

---

<sup>1</sup>Lintner (1965).

<sup>2</sup>Kaufman (1963a, 1963b).

greedy and indicate that optimally these extensions of stochastic programming methods would be applicable to mixed-integer programming problems and not just to linear programs.

As a last possible area of theoretical extension of the present work, there is the whole set of questions concerning decreasing risk aversion in the multiperiod sense. First, one would want to consider the possibility of alternative definitions of decreasing risk aversion when the planning horizon extends over more than one period. Next, for the definition used in this study, one would want to consider further -- despite our lack of success in this study -- the possibility of finding a set of necessary conditions for decreasing risk aversion. Finally, it would also be desirable to derive necessary and/or sufficient conditions for multiperiod decreasing risk aversion when the risks in the several periods are not independent of one another.

The second, and more imperative, direction for further work beyond what has been done here is the actual application of the model and solution procedure that this study has developed. As a first step, the problem of computation raised in Section 5.5 must be considered. Which way of capturing cash-flow interaction among projects is actually best on computational grounds? If it is found that the increment-project approach used in the study is computationally inferior to the method that adds an integer variable (compound project) to take account of interactions, then the procedure presented in Chapter 8 will have to be modified, despite its neat interpretation. To reiterate a "hunch," I doubt that the increment-project approach will, in fact, be found inferior.

Second, there remains the whole question of how best to solve the zero-one subproblem that arises in applying Benders' method. One might consider using one of R. E. Gomory's algorithms<sup>3</sup> or some enumerative scheme that is particularly

---

<sup>3</sup>Gomory (1958, 1960a, 1960b).



well suited to the structure of the zero-one problem as it emerges in Benders' partitioning procedure.<sup>4</sup>

Lastly, there remains the paramount problem of obtaining the several sets of data necessary for application of the proposed model to a particular corporation's problem. First, the multiperiod utility function must be ascertained. Second, one must obtain estimates of the probability distributions of returns from individual projects and estimates of their required cash outlays. Finally, and perhaps most demanding, is the need for estimates of the stochastic and deterministic cash-flow interactions between projects. This is, indeed, a great deal of information for which to ask.

Recent work in several areas, however, permits us at least a glimpse of optimism as to the possibility of meeting these data requirements. First, the work by P. E. Green on determining utility functions implicit in investment decisions made by middle-management personnel gives some ground for hope of obtaining the utility function.<sup>5</sup> Green's study was carried out on a small scale in a chemical corporation and he did find the process of deriving the utility functions a difficult one. Nevertheless, he was able to obtain the functions for the particular level of management personnel he interviewed. Second, the work by D. B. Hertz and his associates involving the derivation of subjective probability distributions for the returns from potential investment projects provides much reason to be hopeful about the possibility of determining such distributions.<sup>6</sup> The tone of Hertz's article is definitely an optimistic one.

To repeat a quotation presented earlier, Hertz writes, "It has been our

---

<sup>4</sup>See Balinski (1965) for a description of the major methods available for solving integer programming problems. Also see Balas (1965) and Glover (1965) for other zero-one programming algorithms. Klevorick (1965a, 1965b) presents the integration of two enumerative schemes for solving zero-one problems into the Benders procedure applied to a special type of plant-location problem.

<sup>5</sup>Green (1961, 1963).

<sup>6</sup>Hertz (1964).

experience that for major capital proposals managements usually make a significant investment in time and funds to pinpoint information about the relevant factors [affecting a project's payoff]. An objective analysis of the values to be assigned to each [factor] can, with little additional effort, yield a subjective probability distribution."<sup>7</sup>

As for the last data requirement, the interactions between the cash flows of different projects, several authors have presented methods for organizing this information that offer some hope of improving the possibility that these data can actually be obtained.<sup>8</sup> Sharpe's "Diagonal Model," for example, which closely resembles Markowitz's suggestion, simplifies the data-gathering and computations in security portfolio problems by assuming that the returns from various securities are related to one another only through their common relationship with some basic factor, for example, the gross national product or the Dow Jones Industrial Index. The return from any security is assumed to be determined only by random elements and by this single basic factor. The basic underlying factor is itself assumed to be a constant plus a random variable. The determination of covariances between the securities is greatly simplified by these assumptions and the amount of information one must obtain about any particular security is also greatly reduced. For example, in an analysis of 100 securities the number of estimates required declines sharply from 5,150 when covariances are determined directly to 302 when the Diagonal Model is used.

Two things ought to be noted about Sharpe's model. First, it can easily be extended to the case where there is more than one underlying index, while still keeping the data-gathering and computational tasks much smaller than if interactions are determined individually. Second, it can easily be transplanted from the realm of portfolio analysis to the area of capital budgeting. In fact,

---

<sup>7</sup>Ibid., p. 100.

<sup>8</sup>Hillier (1964, pp. 58-69, 70-73); Markowitz (1959, pp. 96-101); Sharpe (1963).

Weingartner writes, "Utilization of an index to capture the major effect of covariation seems more appropriate in our [capital-budgeting] context than in portfolio selection. For a single product-line firm, the payoffs from capital expenditures are apt to be related to each other mostly through a variable as total sales."<sup>9</sup>

In the case of each of these data requirements it will, of course, be necessary to educate the decisionmakers whom one is trying to help as to the importance of the needed information. This educational process is a necessity. Only by disciplining the thoughts of decisionmakers as they confront risky investment opportunities can one hope to ensure the improvement of the resulting decisions. With the decisionmakers educated to the importance of the types of data that have just been enumerated, the model and solution procedure of the present study can come into its own as a useful tool in the capital-budgeting process.

---

<sup>9</sup>Weingartner (1966, p. 503).



BIBLIOGRAPHY

I. Books

1. Anderson, T. W., An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, Inc., New York, 1958.
2. Arrow, K. J., Aspects of The Theory of Risk-Bearing, Yrjö Jahnssoonin Saatio, Helsinki, Finland, 1965.
3. Baumol, W. J., Economic Theory and Operations Analysis, Second Edition, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1965.
4. Berle, A. A. and G. C. Means, The Modern Corporation and Private Property, The Macmillan Company, New York, 1932.
5. Bierman, H., Jr. and S. Smidt, The Capital Budgeting Decision; Economic Analysis and Financing of Investment Projects, Second Edition, The Macmillan Company, New York, 1966.
6. Borch, K., The Economics of Uncertainty, Western Management Science Institute, Los Angeles, 1965. [Chapters I-XIII released as Working Papers Nos. 65 (I), 67(II), 69 (III,IV), 70(VII), 72(V), 76 (XI,XII), 77 (XIII), 79 (VI), 82 (VIII, IX), 83 (X)].
7. Boulding, K. and G. Stigler, editors, Readings in Price Theory, Richard D. Irwin, Inc., Chicago, 1952.
8. Bowman, M. J., editor, Proceedings of the Conference on Expectations, Uncertainty and Business Behavior, Social Science Research Council, New York, 1959.
9. Cramér, H., Mathematical Methods of Statistics, Princeton University Press, Princeton, 1946.
10. Cyert, R. M. and J. G. March, A Behavioral Theory of the Firm, Prentice-Hall, Inc. Englewood Cliffs, New Jersey, 1963.
11. Dantzig, G. B., Linear Programming and Extensions, Princeton University Press, Princeton, 1963.
12. Dean, J., Capital Budgeting, Columbia University Press, New York, 1951.
13. Duesenberry, J. S., Income, Saving, and the Theory of Consumer Behavior, Harvard University Press, Cambridge, 1949.
14. Farrar, D. E., The Investment Decision Under Uncertainty, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1962.
15. Feller, W., An Introduction to Probability Theory and Its Applications, John Wiley and Sons, Inc., New York, (Volume I) 1950, (Volume II) 1966.

16. Fisher, I., The Theory of Interest, Macmillan Company, New York, 1930.
17. Gordon, M. J., The Investment, Financing, and Valuation of the Corporation, Richard D. Irwin, Inc., Homewood, Illinois, 1962.
18. Gordon, R. A., Business Leadership in the Large Corporation, University of California Press, Berkeley, 1961.
19. Graves, R. and P. Wolfe, editors, Recent Advances in Mathematical Programming, McGraw-Hill Book Company, New York, 1963.
20. Hadley, G., Linear Algebra, Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1961.
21. Hadley, G., Nonlinear and Dynamic Programming, Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1964.
22. Hakansson, N. H., Optimal Investment and Consumption Strategies for a Class of Utility Functions, Ph.D. Dissertation, University of California, Los Angeles, 1966.
23. Hoel, P. G., Introduction to Mathematical Statistics, Third Edition, John Wiley and Sons, Inc., New York, 1962.
24. Istvan, D. F., The Capital-Expenditure Decision-Making Process in Forty-Eight Large Corporations, D. B. A. Dissertation, School of Business, Indiana University, Bloomington, 1959.
25. Kaufman, G. M., Statistical Decision and Related Techniques in Oil and Gas Exploration, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1963. (b)
26. Kendall, M. G. and A. Stuart, The Advanced Theory of Statistics, Hafner Publishing Company, New York, (Volume 1, Second Edition), 1963.
27. Knight, F. H., Risk, Uncertainty, and Profit, Houghton Mifflin, New York, 1921.
28. Koopmans, T. C., editor, Activity Analysis of Production and Allocation, John Wiley and Sons, Inc., New York, 1951. (a)
29. Luce, R. D. and H. Raiffa, Games and Decisions: Introduction and Critical Survey, John Wiley and Sons, Inc., New York, 1957.
30. Lutz, F. and V. Lutz, The Theory of Investment of the Firm, Princeton University Press, Princeton, 1951.
31. Machinery and Allied Products Institute, MAPI Replacement Manual, Machinery and Allied Products Institute, Chicago, 1950.
32. Marglin, S., Approaches to Dynamic Investment Planning, North-Holland Publishing Company, Amsterdam, 1963.
33. Markowitz, H. M., Portfolio Selection: Efficient Diversification of Investments, John Wiley and Sons, Inc., New York, 1959.

34. Marshall, A., Principles of Economics, Eighth Edition, Macmillan and Company, Ltd., London, 1920.
35. Mason, E. S., editor, The Corporation in Modern Society, Harvard University Press, Cambridge, 1959.
36. Masse, P., Optimal Investment Decisions: Rules for Action and Criteria for Choice, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1962.
37. Mood, A. M. and F. A. Graybill, Introduction to the Theory of Statistics, Second Edition, McGraw-Hill Book Company, Inc., New York, 1963.
38. Muth, J. F. and G. L. Thompson, editors, Industrial Scheduling, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1963.
39. Naslund, B., Decisions Under Risk: Economic Applications of Chance-Constrained Programming, Ph.D. Dissertation, Carnegie Institute of Technology Graduate School of Industrial Administration, Pittsburgh, 1964.
40. Neyman, J., editor, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability; Volume I: Contributions to the Theory of Statistics, University of California Press, Berkeley and Los Angeles, 1961.
41. Olmsted, J. M. H., Advanced Calculus, Appleton-Century Crofts, Inc., New York, 1961.
42. Porterfield, J. T. S., Investment Decisions and Capital Costs, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1965.
43. Robichek, A. A. and S. C. Myers, Optimal Financing Decisions, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1965.
44. Samuelson, P. A., Foundations of Economic Analysis, Harvard University Press, Cambridge, 1947.
45. Savage, L. J., The Foundations of Statistics, John Wiley and Sons, Inc., New York, 1954.
46. Shubik, M., editor, Essays in Mathematical Economics; in Honor of Oskar Morgenstern, Princeton University Press, Princeton, 1967.
47. Solomon, E., editor, The Management of Corporate Capital, The Free Press, Glencoe, Illinois, 1959.
48. Stedry, A. C., Budget Control and Cost Behavior, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1960.
49. Terborgh, G., Dynamic Equipment Policy: A MAPI Study, McGraw-Hill, Inc., New York, 1949.
50. Von Neumann, J. and O. Morgenstern, Theory of Games and Economic Behavior, Second Edition, Princeton University Press, Princeton, 1947.

51. Weingartner, H. M., Mathematical Programming and the Analysis of Capital Budgeting Problems, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1963.
52. Williamson, O. E., The Economics of Discretionary Behavior: Managerial Objectives in a Theory of the Firm, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1964.

## II. Articles

53. Adelson, R. M., "Criteria for Capital Investment: An Approach Through Decision Theory," Operational Research Quarterly, Volume 16, March 1965, pp. 19-50.
54. Archibald, G. C., "Utility, Risk, and Linearity," Journal of Political Economy, Volume LXVII, October 1959, pp. 437-449.
55. Arrow, K. J., "Comments on Duesenberry's 'The Portfolio Approach to the Demand for Money and Other Assets'," The Review of Economics and Statistics, Volume XLV, February 1963 Supplement, pp. 24-27.
56. Arrow, K. J., "The Role of Securities in the Optimal Allocation of Risk-Bearing," Review of Economic Studies, Volume XXXI, April 1964, pp. 91-96.
57. Balas, E., "An Additive Algorithm for Solving Linear Programs With Zero-One Variables," Operations Research, Volume 13, July-August 1965, pp. 517-545.
58. Baldwin, R. H., "How to Assess Investment Proposals," Harvard Business Review, Volume 37, May-June 1959, pp. 98-104.
59. Balinski, M. L., "An Application of Benders Decomposition," Mathematica Working Paper ARO-22, Princeton, November 5, 1963.
60. Balinski, M. L., "Integer Programming: Methods, Uses, Computation," Management Science, Volume 12, October 1965, pp. 253-313.
61. Balinski, M. L. and P. Wolfe, "On Benders Decomposition and A Plant Location Problem," Mathematica Working Paper ARO-24, Princeton, December 27, 1963.
62. Baumol, W. J., "An Expected Gain-Confidence Limit Criterion for Portfolio Selection," Management Science, Volume 10, October 1963, pp. 174-182.
63. Baumol, W. J. and T. Fabian, "Decomposition, Pricing for Decentralization and External Economies," Management Science, Volume 11, September 1964, pp. 1-32.
64. Baumol, W. J. and B. G. Malkiel, "The Firm's Optimal Debt-Equity Combination and the Cost of Capital," Quarterly Journal of Economics, Volume LXXXI, Forthcoming.
65. Baumol, W. J. and R. E. Quandt, "Investment and Discount Rates Under Capital Rationing - A Programming Approach," The Economic Journal, Volume LXXV, June 1965, pp. 317-329.



66. Beale, E. M. L., "A Method of Solving Linear Programming Problems When Some but Not All of the Variables Must Take Integral Values," Statistical Techniques Research Group, Technical Report No. 19, Princeton University, Princeton, July 1958.
67. Beale, E. M. L. and R. E. Small, "Mixed Integer Programming by a Branch and Bound Technique," Proceedings of the International Federation for Information Processing Congress 1965, Volume 2, Sparta Books, Inc., Washington, D. C., pp. 450-451.
68. Benders, J. F., "Partitioning Procedures for Solving Mixed-Variables Programming Problems," Numerische Mathematik, Volume 4, December 1962, pp. 238-252.
69. Bernoulli, D., "Specimen Theoriae Novae de Mensura Sortis," Commentarii Academiae Scientiarum Imperialis Petropolitanae, Volume V, 1738 pp. 175-192. Translated as "Exposition of A New Theory on the Measurement of Risk" by Louise Sommer, Econometrica, Volume 22, January 1954, pp. 23-36.
70. Bodenhorn, D., "On the Problem of Capital Budgeting," Journal of Finance, Volume XIV, December 1959, pp. 473-492.
71. Borch, K., "A Contribution to the Theory of Reinsurance Markets," Skandinavisk Aktuarietidskrift, 1962, pp. 176-189.
72. Borch, K., "A Note on Utility and Attitudes To Risk," Management Science, Volume 9, July 1963, pp. 697-700.
73. Charnes, A. and W. W. Cooper, "Chance-Constrained Programming," Management Science, Volume 6, October 1959, pp. 73-79.
74. Charnes, A. and W. W. Cooper, "Chance Constraints and Normal Deviates," Journal of the American Statistical Association, Volume 57, March 1962, pp. 134-148.
75. Charnes, A. and W. W. Cooper, "Deterministic Equivalents for Optimizing and Satisficing Under Chance Constraints," Operations Research, Volume 11, January-February 1963, pp. 18-39.
76. Charnes, A., W. W. Cooper and M. H. Miller, "Application of Linear Programming to Financial Budgeting and the Costing of Funds," Journal of Business, Volume XXXII, January 1959, pp. 20-46. Reprinted in E. Solomon (1959), pp. 229-255. Page references in text are all to E. Solomon (1959).
77. Cord, J., "A Method for Allocating Funds to Investment Projects When Returns Are Subject to Uncertainty," Management Science, Volume 10, January 1964, pp. 335-341.
78. Dantzig, G. B., "Linear Programming Under Uncertainty," Management Science, Volume 1, April and July 1955, pp. 197-206.
79. Dantzig, G. B. and A. Madansky, "On the Solution of Two-Stage Linear Programs Under Uncertainty," in J. Neyman (1961), pp. 165-176.

80. Dantzig, G. B. and P. Wolfe, "A Decomposition Principle for Linear Programs," Operations Research, Volume 8, January-February 1960, pp. 101-111.
81. Dean, J., "Measuring the Productivity of Capital," Harvard Business Review, Volume 32, January-February 1954, pp. 120-130. Reprinted in E. Solomon (1959), pp. 21-34. Page references in text are all to E. Solomon (1959).
82. Dean, J. and W. Smith, "Has MAPI a Place in a Comprehensive System of Capital Controls?" Journal of Business, Volume XXVIII, October 1955, pp. 261-274. Reprinted in E. Solomon (1959), pp. 293-306. Page references in text are all to E. Solomon (1959).
83. Driebeek, N. J., "An Algorithm for the Solution of Mixed Integer Programming Problems," Management Science, Volume 7, March 1966, pp. 576-587.
84. Durand, D., "Costs of Debt and Equity Funds for Business: Trends and Problems of Measurement," in Conference on Research in Business Finance, National Bureau of Economic Research, New York, 1952, pp. 215-247. Reprinted in E. Solomon (1959), pp. 91-127. Page references in text are all to E. Solomon (1959).
85. Dyckman, T. R., "Allocating Funds to Investment Projects When Returns Are Subject to Uncertainty - A Comment," Management Science, Volume 11, November 1964, pp. 348-350.
86. Elmaghraby, S. E., "Allocation Under Uncertainty When the Demand has Continuous Distribution Function," Management Science, Volume 6, April 1960, pp. 270-294.
87. Elmaghraby, S. E., "An Approach to Linear Programming Under Uncertainty," Operations Research, Volume 7, March-April 1959, pp. 208-216.
88. Ferguson, A. R. and G. B. Dantzig, "The Allocation of Aircraft to Routes: An Example of Linear Programming Under Uncertain Demand," Management Science, Volume 3, October 1956, pp. 45-73.
89. Freund, R. J., "The Introduction of Risk into a Programming Model," Econometrica, Volume 24, July 1956, pp. 253-263.
90. Friedman, M. and L. J. Savage, "The Utility Analysis of Choices Involving Risk," Journal of Political Economy, Volume LVI, August 1948, pp. 279-304. Reprinted in K. Boulding and G. Stigler (1952), pp. 57-96. Page references in text are all to K. Boulding and G. Stigler (1952).
91. Gass, S. I. and T. L. Saaty, "The Computational Algorithm for the Parametric Objective Function," Naval Research Logistics Quarterly, Volume 2, March-June 1955, pp. 39-45. (a)
92. Gass, S. I. and T. L. Saaty, "The Parametric Objective Function, Part I," Operations Research, Volume 2, August 1954, pp. 316-319.
93. Gass, S. I. and T. L. Saaty, "The Parametric Objective Function, Part II - Generalization," Operations Research, Volume 3, November 1955, pp. 395-401. (b)

94. Glover, F., "A Multiphase - Dual Algorithm for the Zero-One Integer Programming Problem," Operations Research, Volume 13, November-December 1965, pp. 879-919.
95. Goldman, S. M. and H. Uzawa, "A Note on Separability in Demand Analysis," Econometrica, Volume 32, July 1964, pp. 387-398.
96. Gomory, R. E., "All-Integer Integer Programming Algorithm," in J. F. Muth and G. L. Thompson (1963), pp. 193-206. First issued as IBM Research Center Research Report RC-189, Yorktown Heights, New York, January 29, 1960. (a)
97. Gomory, R. E., "An Algorithm for Integer Solutions to Linear Programs," in R. Graves and P. Wolfe (1963), pp. 269-302. First issued as Princeton-IBM Mathematics Research Project Technical Report No. 1, Princeton, November 17, 1958.
98. Gomory, R. E., "An Algorithm for the Mixed Integer Problem," Rand Corporation RM-2597, Santa Monica, California, July 7, 1960. (b).
99. Gordon, M. J. "The Payoff Period and The Rate of Profit," Journal of Business, Volume XXVIII, October 1955, pp. 253-260. Reprinted in E. Solomon (1959), pp. 48-55. Page references in text are all to E. Solomon (1959).
100. Gordon, M. J. and E. Shapiro, "Capital Equipment Analysis: The Required Rate of Profit," Management Science, Volume 3, October 1956, pp. 102-110. Reprinted in E. Solomon (1959), pp. 141-149. Page references in text are all to E. Solomon (1959).
101. Gort, M., "The Planning of Investment: A Study of Capital Budgeting in the Electric Power Industry," Journal of Business, Part I, Volume XXIV, April 1951, pp. 79-95; Part II, Volume XXIV, July 1951, pp. 181-202.
102. Green, P. E., "The Derivation of Utility Functions in a Large Industrial Firm," paper given at the First Joint National Meeting of the Operations Research Society of America and the Institute of Management Sciences, 1961.
103. Green, P. E., "Risk Attitudes and Chemical Investment Decisions," Chemical Engineering Progress, Volume 59, January 1963, pp. 35-40.
104. Harris, P. M. J., "An Algorithm for Solving Mixed Integer Linear Programmes," Operational Research Quarterly, Volume 15, June 1964, pp. 117-132.
105. Heller, W., "The Anatomy of Investment Decisions," Harvard Business Review, Volume 29, March 1951, pp. 95-103.
106. Hertz, D. B., "Risk Analysis in Capital Investment," Harvard Business Review, Volume 42, January-February 1964, pp. 95-106.
107. Hetrick, J. C., "Mathematical Models in Capital Budgeting," Harvard Business Review, Volume 39, January-February 1961, pp. 49-64.
108. Hill, H. G., Jr., "Capital Expenditure Management," Journal of Business, Volume XXVIII, October 1955, pp. 285-290. Reprinted in E. Solomon (1959), pp. 282-287. Page references in text are all to E. Solomon (1959).

109. Hillier, F.S., "Derivation of Probabilistic Information for the Evaluation of Risky Investments," Management Science, Volume 9, April 1963, pp. 443-457.
110. Hillier, F. S., "The Evaluation of Risky Interrelated Investments," Department of Statistics, Technical Report No. 73, Stanford University, Stanford, California, July 24, 1964.
111. Hirshleifer, J., "Efficient Allocation of Capital in an Uncertain World," American Economic Review, Volume LIV, May 1964, pp. 77-85.
112. Hirshleifer, J., "Investment Decision Under Uncertainty: Applications of the State-Preference Approach," Quarterly Journal of Economics, Volume LXXX, May 1966, pp. 252-277.
113. Hirshleifer, J., "Investment Decision Under Uncertainty: Choice-Theoretic Approaches," Quarterly Journal of Economics, Volume LXXIX, November 1965, pp. 509-536.
114. Hirshleifer, J., "On the Theory of Optimal Investment Decision," Journal of Political Economy, Volume LXVI, August 1958, pp. 329-352. Reprinted in E. Solomon (1959), pp. 205-228. Page references in text are all to E. Solomon (1959).
115. Hirshleifer, J., "Risk, the Discount Rate, and Investment Decisions," American Economic Review, Volume LI, May 1961, pp. 112-120.
116. Istvan, D. F., "The Economic Evaluation of Capital Expenditures," Journal of Business, Volume XXXIV, January 1961, pp. 45-51.
117. Kaufman, G. M., "Sequential Investment Analysis Under Uncertainty," Journal of Business, Volume XXXVI, January 1963, pp. 39-64. (a)
118. Kaysen, C., "The Corporation: How Much Power? What Scope?" in E. S. Mason (1959), pp. 85-105.
119. Kelley, J. E., "The Cutting-Plane Method for Solving Convex Programs," Journal of the Society of Industrial and Applied Mathematics, Volume 8, December 1960, pp. 703-712.
120. Klevorick, A. K., "An Algorithm for the Plant Location Problem: An Application of Benders' Decomposition," Mathematica Working Paper ARO-32, Princeton, September 1, 1965. (a)
121. Klevorick, A. K., "An Algorithm for the Plant Location Problem: A Second Application of Benders' Decomposition," Mathematica Working Paper ARO-33, Princeton, September 15, 1965. (b)
122. Koopmans, T. C., "Analysis of Production as an Efficient Combination of Activities," in T. C. Koopmans (1951a), pp. 33-97. (b)
123. Kuh, E., "Capital Theory and Capital Budgeting," Metroeconomica, Volume XII, August-December 1960, pp. 64-80.

124. Kunzi, H. P. and W. Oettli, "Integer Quadratic Programming," in R. Graves and P. Wolfe (1963), pp. 303-308.
125. Land, A. H. and A. G. Doig, "An Automatic Method of Solving Discrete Programming Problems," Econometrica, Volume 28, July 1960, pp. 497-520.
126. Latané, H. A., "Criteria for Choice Among Risky Ventures," Journal of Political Economy, Volume LXVII, April 1959, pp.144-155.
127. Lintner, J., "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," The Review of Economics and Statistics, Volume XLVII, February 1965, pp. 13-37.
128. Lorie, J. H. and L. J. Savage, "Three Problems in Rationing Capital," Journal of Business, Volume XXVIII, October 1955, pp. 229-239. Reprinted in E. Solomon (1959), pp. 56-66. Page references in text are all to E. Solomon (1959).
129. McLean, J. G., "How to Evaluate New Capital Investments," Harvard Business Review, Volume 36, November-December 1958, pp. 59-69.
130. Madansky, A., "Linear Programming Under Uncertainty," in R. Graves and P. Wolfe (1963), pp. 103-110.
131. Madansky, A., "Methods of Solutions of Linear Programs under Uncertainty," Operations Research, Volume 10, July-August 1962, pp. 463-471.
132. Manne, A. S., "Optimal Dividend and Investment Policies for A Self-Financing Business Enterprise," Graduate School of Business Working Paper No. 93, Stanford Graduate School of Business, Stanford, California, February 4, 1966.
133. Markowitz, H. M., "The Utility of Wealth," Journal of Political Economy, Volume LX, April 1952, pp. 151-158.
134. Marschak, J., "Rational Behavior, Uncertain Prospects, and Measurable Utility," Econometrica, Volume 18, April 1950, pp. 111-141.
135. Marschak, T., "Capital Budgeting and Pricing in the French Nationalized Industries," Journal of Business, Volume XXXIII, April 1960, pp. 133-156.
136. Martin, A. D., "Mathematical Programming of Portfolio Selection," Management Science, Volume I, January 1955, pp. 152-166.
137. Menger, K., "The Role of Uncertainty in Economics," translated from Zeitschrift für Nationalökonomie, Band V, 1934, pp. 459-485. Reprinted in M. Shubik (1967), pp. 211-231. Page references in text are all to M. Shubik (1967).
138. Miller, M. H. and F. Modigliani, "Some Estimates of the Cost of Capital to the Electric Utility Industry, 1954-57," American Economic Review, Volume LVI, June 1966, pp. 333-391.
139. Modigliani, F. and K. J. Cohen, "The Significance and Uses of Ex Ante Data - A Summary View," in M. J. Bowman (1959), pp. 151-164.

140. Modigliani, F. and M. H. Miller, "The Cost of Capital, Corporation Finance and the Theory of Investment," American Economic Review, Volume XLVIII, June 1958, pp. 261-297. Reprinted in E. Solomon (1959), pp. 150-181. Page references in text are all to E. Solomon (1959).
141. Naslund, B., "A Model of Capital Budgeting Under Risk," Journal of Business, Volume XXXIX, April 1966, pp. 257-271.
142. Naslund, B. and A. Whinston, "Model of Decision Making Under Risk," Metroeconomica, Volume XVI, May-August 1964, pp. 81-94.
143. Naslund, B. and A. Whinston, "A Model of Multi-Period Investment Under Uncertainty," Management Science, Volume 8, January 1962, pp. 184-200.
144. Norton, F. E., "Administrative Organization in Capital Budgeting," Journal of Business, Volume XXVIII, October 1955, pp. 291-295. Reprinted in E. Solomon (1959), pp. 288-292. Page references in text are all to E. Solomon (1959).
145. Pratt, J. W., "Risk Aversion in the Small and in the Large," Econometrica, Volume 32, January-April 1964, pp. 122-136.
146. Pullara, S. J. and L. R. Walker, "The Evaluation of Capital Expenditure Proposals: A Survey of Firms in the Chemical Industry," The Journal of Business, Volume XXXVIII, October 1965, pp. 403-408.
147. Pye, G. B., "Portfolio Selection and Security Prices," The Review of Economics and Statistics, Volume XLIX, February 1967, pp. 111-115.
148. Pye, G. B., "Present Values for Imperfect Capital Markets," Journal of Business, Volume XXXIX, Part 1, January 1966, pp. 45-51.
149. Reiter, S., "Choosing An Investment Program Among Interdependent Projects," Review of Economic Studies, Volume XXX, January 1963, pp. 32-36.
150. Reiter, S. and D. B. Rice, "Discrete Optimizing Solution Procedures for Linear and Nonlinear Integer Programming Problems," Management Science, Volume 12, July 1966, pp. 829-850.
151. Reiter, S. and G. R. Sherman, "Allocating Indivisible Resources Affording External Economies or Diseconomies," International Economic Review, Volume 3, January 1962, pp. 108-135.
152. Reiter, S. and G. R. Sherman, "Discrete Optimizing," Journal of the Society of Industrial and Applied Mathematics, Volume 13, September 1965, pp. 864-889.
153. Renshaw, E., "A Note on the Arithmetic of Capital Budgeting Decisions," Journal of Business, Volume XXX, July 1957, pp. 193-201. Reprinted in E. Solomon (1959), pp. 80-88. Page references in text are all to E. Solomon (1959).
154. Roberts, H. V., "Current Problems in the Economics of Capital Budgeting," Journal of Business, Volume XXX, January 1957, pp. 12-16. Reprinted in E. Solomon (1959), pp. 198-202. Page references in text are all to E. Solomon (1959).

155. Roy, A. D., "Safety First and the Holding of Assets," Econometrica, Volume 20, July 1952, pp. 431-449.
156. Russell, W. R. and P. E. Smith, "A Comment on Baumol (E,L) Efficient Portfolios," Management Science, Volume 12, March 1966, pp. 619-621.
157. Sharpe, W. F., "A Simplified Model of Portfolio Selection," Management Science, Volume 9, January 1963, pp. 277-293.
158. Shillinglaw, G., "Residual Values in Investment Analysis," Journal of Business, Volume XXVIII, October 1955, pp. 275-284. Reprinted in E. Solomon (1959), pp. 259-268. Page references in text are all to E. Solomon (1959).
159. Solomon, E., "The Arithmetic of Capital-Budgeting Decisions," Journal of Business, Volume XXIX, April 1956, pp. 124-129. Reprinted in E. Solomon (1959), pp. 74-79. Page references in text are all to E. Solomon (1959).
160. Solomon, E., "Measuring A Company's Cost of Capital," Journal of Business, Volume XXVIII, October 1955, pp. 240-252. Reprinted in E. Solomon (1959), pp. 128-140. Page references in text are all to E. Solomon (1959).
161. Steindl, J., "On Risk," Oxford Economic Papers, Volume 5, June 1941, pp. 43-53.
162. Teichroew, D., A. A. Robichek and M. Montalbano, "An Analysis of Criteria for Investment and Financing Decisions Under Certainty," Management Science, Volume 12, November 1965, pp. 151-179.
163. Terborgh, G., "Some Comments on the Dean-Smith Article on the MAPI Formula," Journal of Business, Volume XXIX, April 1956, pp. 138-140. Reprinted in E. Solomon (1959), pp. 307-309. Page references in text are all to E. Solomon (1959).
164. Thompson, G. L., W. W. Cooper and A. Charnes, "Characterization by Chance-Constrained Programming," in R. Graves and P. Wolfe (1963), pp. 113-130.
165. Tobin, J., "Liquidity Preference As Behavior Toward Risk," Review of Economic Studies, Volume XXVI, February 1958, pp. 65-86.
166. Turvey, R., "Present Value versus Internal Rate of Return— An Essay in the Theory of Third Best," The Economic Journal, Volume LXXIII, March 1963, pp. 93-98.
167. Van Moeseke, P., "Stochastic Linear Programming," Yale Economic Essays, Volume 5, Spring 1965, pp. 197-253.
168. Weingartner, H. M., "Capital Budgeting of Interrelated Projects: Survey and Synthesis," Management Science, Volume 12, March 1966, pp. 485-516.
169. Williams, A. C., "Approximation Formulas for Stochastic Linear Programming," Journal of the Society of Industrial and Applied Mathematics, Volume 14, July 1966, pp. 668-677.

170. Williams, A. C., "On Stochastic Linear Programming," Journal of the Society of Industrial and Applied Mathematics, Volume 13, September 1965, pp. 927-940.
171. Williams, A. C. and J. I. Nassar, "Financial Measurement of Capital Investments," Management Science, Volume 12, July 1966, pp. 851-864.
172. Williamson, O. E., "Managerial Discretion and Business Behavior," American Economic Review, Volume LIII, December 1963, pp. 1032-1057. (b)
173. Williamson, O. E., "A Model of Rational Managerial Behavior," in R. M. Cyert and J. G. March (1963), pp. 237-252. (a)
174. Witzgall, C., "An All-Integer Programming Algorithm With Parabolic Constraints," Journal of the Society of Industrial and Applied Mathematics, Volume XI, December 1963, pp. 855-871.
175. Yaari, M. E., "Convexity in the Theory of Choice Under Risk," Quarterly Journal of Economics, Volume LXXIX, May 1965, pp. 278-290.
176. Zangwill, W. I., "Media Selection by Decision Programming," Journal of Advertising Research, Volume V, September 1965, pp. 30-36.



ABSTRACT

This study considers the problem of capital budgeting under risk for a firm operating in an imperfect capital market. The "capital-budgeting problem" is defined as the problem of allocating fixed budget dollars in each of several time periods among competing investment proposals. The total fixed amount of money available for investment purposes in any period and the composition of the financing of that amount are taken as given. Subject to an absolute borrowing limit in each period, the firm can borrow and lend funds at constant but divergent rates of interest with the borrowing rate exceeding the lending rate. The gross returns from the potential projects are stochastic; everything else is assumed to be known with certainty.

The study begins with a general discussion of decisionmaking over time in a risk environment. Attitudes toward risk in a multiperiod context, in particular, risk aversion and decreasing risk aversion, are characterized in terms of properties of the decisionmaker's multiperiod utility function. The new concept of risk balancing over time is introduced.

Temporarily assuming away risk, the goal of the investment program of the firm is considered. It is argued that the appropriate objective is maximization of a nonlinear utility function that is management's perception of the utility the owners derive from their consumption alternatives over the firm's T-period planning horizon. Since the true arguments of this function are not measurable, the amounts available for withdrawal from the firm in the T periods are taken as proxy measures of the arguments' values. Previously suggested approaches for investment-project selection are shown to fall short of optimizing this objective function. A new nonlinear programming model is then presented. The model is closely akin to previous programming approaches. It has, however, several distinctive features, namely, (1) a nonlinear utility function as the maximand,

(2) a particular period-analytic view of the firm's cash flows, (3) special attention to cash-flow interactions among projects, and (4) a particular combination of capital-market opportunities and disbursement activities.

Before reconsidering this programming model in light of the risks assumed to exist, previous suggestions for budgeting capital in this environment are evaluated. All of them, including the popular mean-variance approach, are found to be unsatisfactory. Taking the expected-utility theorem as the basis for rational behavior under risk, a set of axioms that a multiperiod utility function serving as the capital-budgeting objective function ought to satisfy is presented. The utility function, it is asserted, must possess positive marginal utility, be bounded, show risk aversion and decreasing risk aversion, and obviate the need to specify fully the joint frequency function of withdrawals from the firm. A cubic utility function, separable or nonseparable, is shown to be the lowest-degree polynomial utility function satisfying these axioms. The decisionmaker applying the expected-utility theorem to a cubic utility function satisfying the axioms is shown to prefer ceteris paribus increases in expected return and in positive skewness and ceteris paribus decreases in variance.

The assumptions about the risk environment ensure that the constraint set faced by the firm under risk is the same as under certainty. The objective function that results from the axiomatic investigation is therefore combined with the constraint set of the certainty model to yield a programming model of capital budgeting in the risk and imperfect-capital-market environment described. The resulting model is a linear mixed-integer program. A solution procedure based upon J. F. Benders' partitioning algorithm is presented for solving the resulting capital-budgeting model. The algorithm's progress toward a solution is lent an economic interpretation.

**DOCUMENT CONTROL DATA - R&D**

*(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)*

1. ORIGINATING ACTIVITY (Corporate author)

Princeton University

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

3. REPORT TITLE

CAPITAL BUDGETING UNDER RISK: A MATHEMATICAL-PROGRAMMING APPROACH

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Research Memorandum No. 89, Sept. 1967

5. AUTHOR(S) (Last name, first name, initial)

Alvin Keith Klevorick

6. REPORT DATE

September 1967

7a. TOTAL NO. OF PAGES

365

7b. NO. OF REFS

176

8a. CONTRACT OR GRANT NO.

Nonr. 1858(16)

b. PROJECT NO.

c. Task No. NR 047-086

d.

9a. ORIGINATOR'S REPORT NUMBER(S)

Research Memorandum No. 89

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

Logistics and Mathematical Branch  
Office of Naval Research  
Washington, D.C. 20360

13. ABSTRACT

This study considers the problem of capital budgeting under risk for a firm operating in an imperfect capital market. The "capital-budgeting problem" is defined as the problem of allocating fixed budget dollars in each of several time periods among competing investment proposals. The total fixed amount of money available for investment purposes in any period and the composition of the financing of that amount are taken as given. Subject to an absolute borrowing limit in each period, the firm can borrow and lend funds at constant but divergent rates of interest with the borrowing rate exceeding the lending rate. The gross returns from the potential projects are stochastic; everything else is assumed to be known with certainty.

The study begins with a general discussion of decisionmaking over time in a risk environment. Attitudes toward risk in a multiperiod context, in particular, risk aversion and decreasing risk aversion, are characterized in terms of properties of the decisionmaker's multiperiod utility function. The new concept of risk balancing over time is introduced.

Temporarily assuming away risk, the goal of the investment program of the firm is considered. It is argued that the appropriate objective is maximization of a nonlinear utility function that is management's perception of the utility the owners derive from their consumption alternatives over the firm's T-period planning horizon. Since the true arguments of this function are not measurable, the amounts available for withdrawal from the firm in the T periods are taken as proxy measures of the

Security Classification

| 14. KEY WORDS   | LINK A |    | LINK B |    | LINK C |    |
|---|--------|----|--------|----|--------|----|
|   | ROLE   | WT | ROLE   | WT | ROLE   | WT |
| capital budgeting under risk<br>imperfect capital market<br>decisionmaking over time in a risk environment<br>attitudes toward risk in a multi-period context<br>risk aversion decreasing risk aversion<br>multiperiod utility function<br>risk balancing over time<br>goal of the investment program of the firm<br>maximization of a nonlinear utility function<br>management's perception of the owners' utility<br>stockholders' consumption possibilities<br>nonlinear programming approach<br>interactions among projects' cash flows<br>artificial-project approach to cash-flow interactions<br>(con't on sheet attached) |        |    |        |    |        |    |

INSTRUCTIONS

1. **ORIGINATING ACTIVITY:** Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (*corporate author*) issuing the report.

2a. **REPORT SECURITY CLASSIFICATION:** Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. **GROUP:** Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. **REPORT TITLE:** Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. **DESCRIPTIVE NOTES:** If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. **AUTHOR(S):** Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. **REPORT DATE:** Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. **TOTAL NUMBER OF PAGES:** The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. **NUMBER OF REFERENCES:** Enter the total number of references cited in the report.

8a. **CONTRACT OR GRANT NUMBER:** If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. **PROJECT NUMBER:** Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. **ORIGINATOR'S REPORT NUMBER(S):** Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. **OTHER REPORT NUMBER(S):** If the report has been assigned any other report numbers (*either by the originator or by the sponsor*), also enter this number(s).

10. **AVAILABILITY/LIMITATION NOTICES:** Enter any limitations on further dissemination of the report, other than those

imposed by security classification, using standard statements such as:

- (1) "Qualified requesters may obtain copies of this report from DDC."
- (2) "Foreign announcement and dissemination of this report by DDC is not authorized."
- (3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through \_\_\_\_\_."
- (4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through \_\_\_\_\_."
- (5) "All distribution of this report is controlled. Qualified DDC users shall request through \_\_\_\_\_."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. **SUPPLEMENTARY NOTES:** Use for additional explanatory notes.

12. **SPONSORING MILITARY ACTIVITY:** Enter the name of the departmental project office or laboratory sponsoring (*paying for*) the research and development. Include address.

13. **ABSTRACT:** Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (TS), (S), (C), or (U).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. **KEY WORDS:** Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

## ABSTRACT (continued)

arguments' values. Previously suggested approaches for investment-project selection are shown to fall short of optimizing this objective function. A new nonlinear programming model is then presented. The model is closely akin to previous programming approaches. It has, however, several distinctive features, namely, (1) a nonlinear utility function as the maximand, (2) a particular period-analytic view of the firm's cash flows, (3) special attention to cash-flow interactions among projects, and (4) a particular combination of capital-market opportunities and disbursement activities.

Before reconsidering this programming model in light of the risks assumed to exist, previous suggestions for budgeting capital in this environment are evaluated. All of them, including the popular mean-variance approach, are found to be unsatisfactory. Taking the expected-utility theorem as the basis for rational behavior under risk, a set of axioms that a multiperiod utility function serving as the capital-budgeting objective function ought to satisfy is presented. The utility function, it is asserted, must possess positive marginal utility, be bounded, show risk aversion and decreasing risk aversion, and obviate the need to specify fully the joint frequency function of withdrawals from the firm. A cubic utility function, separable or nonseparable, is shown to be the lowest-degree polynomial utility function satisfying these axioms. The decisionmaker applying the expected-utility theorem to a cubic utility function satisfying the axioms is shown to prefer ceteris paribus increases in expected return and in positive skewness and ceteris paribus decreases in variance.

The assumptions about the risk environment ensure that the constraint set faced by the firm under risk is the same as under certainty. The objective function that results from the axiomatic investigation is therefore combined with the constraint set of the certainty model to yield a programming model of capital budgeting in the risk and imperfect-capital-market environment described. The resulting model is a linear mixed-integer program. A solution procedure based upon J. F. Benders' partitioning algorithm is presented for solving the resulting capital-budgeting model. The algorithm's progress toward a solution is lent an economic interpretation.

## KEY WORDS (continued)

risk environment  
chance-constrained programming  
mean-variance approach  
expected-utility hypothesis  
axiomatic approach  
cubic utility function  
interperiod interactions  
skewness  
linear-mixed-integer program  
Bender's partitioning algorithm for mixed-variables problems  
economic interpretation