MI

AE REPORT # 390

Agricultural Economics Report

REPORT NO. 390 MARCH 1981

THEORETICAL AND PRACTICAL MODELS FOR INVESTMENT AND DISINVESTMENT DECISION MAKING UNDER UNCERTAINTY IN THE ENERGY SUPPLY INDUSTRY

Edited by:

Lindon J. Robison
Michael H. Abkin

Department of
Agricultural Economics
MICHIGAN STATE
UNIVERSITY
East Lansing

THEORETICAL AND PRACTICAL MODELS FOR INVESTMENT AND DISINVESTMENT DECISION MAKING UNDER UNCERTAINTY IN THE ENERGY SUPPLY INDUSTRY

Edited by

Lindon J. Robison and Michael H. Abkin

A Final Report to the
Electric Power Research Institute

Project No. RP 1220-2
December 1980

Abstract

The report reviews, extends and adapts for empirical work the theoretical framework for decisions regarding the investment, disinvestment and use of durable assets under conditions of uncertainty. The relevance to the utility industry is that their ability to supply a time-varying demand for their product depends on the accumulation of investments, disinvestments and use of durable assets.

In this report, a general framework is presented for prescribing durable investment and related decisions. In order to do so, relationships are identified that account for capacity and inventory costs as well as identifying the interdependencies between the two in a dynamic (time) analysis.

To aid in the decisionmaking process under uncertainty, while accounting for unique risk preferences, a new risk-efficiency criterion is reviewed. In addition, a measurement technique for identifying risk preferences is introduced.

Utilizing the theoretical developments and the new risk-efficiency criterion for ordering action choices, a general simulation model was then developed. The model, of intermediate complexity, is intended to be used in the initial stages of screening investment options. In an empirical test, the newly developed model obtained nearly the same results as the well-established Minimum Revenue Requirements method. The new theory and simulation model, however, have broader capabilities for application at the capacity or corporate planning level.

Agricultural Economics
Report #390
March 1981

THEORETICAL AND PRACTICAL MODELS FOR INVESTMENT AND DISINVESTMENT DECISION MAKING UNDER UNCERTAINTY IN THE ENERGY SUPPLY INDUSTRY

Edited by

Lindon J. Robison and Michael H. Abkin

A Final Report to the
Electric Power Research Institute

Project No. RP 1220-2

March 1981

Department of Agricultural Economics
Michigan State University
East Lansing, Michigan 48824

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF TABLES

ACKNOWLEDGEMENTS

The authors of this report have benefited from reviews and discussions with several individuals. First, Dr. A. N. Halter of EPRI provided continuous support during the project, and early discussions with him helped focus our direction. Since the results of this project were intended to have applied relevance to the energy industry, it was critical that we received direct and continuous feedback from a utility representative. James Parker, of Consumers Power Company of Jackson, Michigan, provided such support. In all phases of this project, he freely interacted with members of the research team.

In March of 1980, an interim report conference was held on the campus of Michigan State University. At that conference, we presented our findings, and experts from around the country were asked to provide formal responses. We thank the following for attending and providing input at a critical stage of the project: Dr. Peter Barry, University of Illinois; Dr. J. Roy Black, Michigan State University; Ron Calcatera, Consumers Power Company; Dr. Albert N. Halter, EPRI; Dr. James Jonish, Texas Tech University; Dr. Otto Krauss, Michigan State University; Dr. Norman Obst, Michigan State University; Dr. Gerald Park, Michigan State University; James Parker, Consumers Power Company; and Ron Radke, Michigan Public Service Commission.

Many others reviewed this report at various stages, including Mark Cochran and Dr. Gerald Schwab. We also express our sincere thanks to our typists, Jeanette Barbour, Julia McKay, Debbie Andrews and Linda Todd, who, against insurmountable obstacles, transformed our writings into a professionally prepared manuscript.

The Michigan Agricultural Experiment Station also provided us support throughout this project, particularly in this report-writing phase.

Finally, we assume complete responsibility for the contents of this report and for the errors which, despite the best efforts of our reviewers, likely remain.

Michigan Agricultural Experiment Station Publication No. 9927.

EXECUTIVE SUMMARY

Introduction and Organization of the Report

Those industries which supply energy in various forms have characteristics which make them somewhat unique. This report originated because of an interest in analyzing two of those characteristics: (1) the capital-intensive nature of the industry; and (2) the uncertainty of the environment in which the industries operate. These characteristics have important implications for the industries' ability to alter their supply of energy in response to changes in economic conditions. The primary objective of this study is to improve analysts' ability to describe how energy supply industries respond to changes in their environment.

In order to achieve this objective, the work was divided into four tasks. Task 1, reported in Chapters II and III, reviewed selected previous studies on the causes of supply variations in the major energy-producing industries in order to obtain, where possible, empirical measures that would aid in determining the relative importance of variables producing energy supply variations. In addition to literature on the empirical importance of supply variations, selected theoretical models are also reviewed in Chapter II. Finally, the review in Chapter III of the Department of Energy's Mid-Range Energy Forecasting System (MREFS) supplements the reviews of the micro models described in Chapter II.

The review in Task 1 points out the need to extend the existing theory of investment, disinvestment, and use of durable assets and combine it with a theory of decision making under uncertainty--the subject of Task 2.

Task 2, reported in Chapters IV and V, extended the existing investment/disinvestment theory and included the impacts of uncertainty on the decision process. In meeting the requirements of Task 2, this project recognized two different areas of uncertainty analysis, namely, the application of uncertainty to economic models and the development of criteria for ordering action choices.

Task 3, reported in Chapter VI of this report, developed an empirical model that could be used to implement the theoretical model and decision tools presented in Chapters IV and V. The model is one of intermediate complexity and is intended primarily as a preliminary planning tool. A user's guide, documenting the model and software package, has been submitted to EPRI as a separate report.

One way to evaluate new modeling approaches is to compare a new model's performance with an existing model. Chapter VII reports such a comparison, as called for in the project's last task, Task 4. The comparison involves a 1977 decision faced by Consumers Power Company of Jackson, Michigan, of whether to reactivate hydro stations along the Boardman River near Traverse City, Michigan. Their evaluation of this project was based on the widely used and familiar minimum revenue requirements method. To test the theoretical developments and usefulness of the model developed in this report, we employed the newly developed model using Consumers Power data and compared the results to those of their analysis using the minimum revenue requirements method.

The report concludes, in Chapter VIII, with recommendations for future work.

A Review of the Literature

Because of the capital intensity of the energy supply industry, decisions regarding the acquisition, use, and disposal of durable assets are of major importance for managers of energy supply firms. Decisions concerning the acquisition and/or disposal of durable assets are inherently different from decisions regarding the acquisition of non-durable assets. Durable assets are typically available in discrete units and are capable of being used during more than one production period. Thus, decisions regarding the acquisition, use, and disposal of durable assets require information about future production periods. This is not the case for non-durable assets, since they are entirely used up in our production period.

Minimum Revenue Requirements Method

Much of the economic literature on the theory of investment and disinvestment decisions is reviewed in Baquet (1977a). The theory had its origin in Marshall's Principles of Economics. The basic investment decision rule advanced by Marshall still holds: balance current costs against discounted future gratifications. The rule is still applied today and appears as valid as ever. The practical application of Marshall's rule in the utility industry is Jeynes' minimum revenue requirements discipline (MRRD).

According to Jeynes, the primary objective of the utility firm is to discover the alternative investment plans which will result in the minimum outlays of capital investments and periodic expenses throughout the service life of the durables. The minimum revenue requirements are defined to be the revenues which must be obtained in order to cover all expenses incurred, including the company's minimum acceptable return on investors' capital (Jeynes, p. 62). The difference between minimum revenue requirements, on the one hand, and periodic expenses and returns on capital investments, on the other, is defined as the profit incentive.

If a project is to be undertaken, its earnings must be greater than the cost of financing, which is to be minimized. So, the fundamental objective is to determine the minimum average return as a percentage of investors' committed capital.

The MRRD method provides a framework for making capital investment decisions. However, within that framework, some important aspects of investment decisions are ignored. For example, the impact of uncertainty is not addressed explicitly. Moreover, there is no framework for determining the optimal life or utilization rates of the durables.

Recent theoretical work by Baquet (1978) does not rely on the fixed usage assumption. Moreover, it permits the durable to be used at different rates over its productive life. While this is a more realistic approach, it complicates the calculation of the durable's value in use since the value in use is now more closely tied to future production decisions. In this framework, the durable's productive life is no longer fixed but variable.

The importance of accounting for varying rates of production in the analysis lies not so much in a need to determine supply, since the supply is required to meet demand, but in the need to know how to adjust service extraction rates for available durable units to meet a time-varying demand.

Effects of Asset Fixity on Supply Response Models

Economic theorists have long recognized the limits that fixed factors of production place on a firm's ability to respond to changes in its economic environment. Early theoreticians, however, considered assets to be fixed in a physical sense. It was not until the 1950s that Johnson, Willet, Hardin, and others recognized that differences between acquisition and salvage values could fix an asset to the firm in an economic sense. Edwards wrote, "an asset is fixed if it isn't worth varying," suggesting an asset is fixed if its marginal value in use is bounded above by its acquisition price and below by its salvage price.

Johnson (1960) saw the supply response implications of acquisition/salvage price differentials. In short, they suggest a different supply function for increasing output than for decreasing output, since underinvestment decisions can be more easily corrected (by additional investment) than can overinvestment decisions because overinvestments often result in assets being fixed.

While Johnson laid the theoretical groundwork for incorporating asset fixity in supply response work, there were some aspects he recognized but did not develop. To simplify the stock/flow conversion problem, he assumed services are extracted from the durable at a fixed rate. Idachaba made an effort to remove this theoretical limitation. Then Baquet made additional refinements by addressing a specific production problem with services provided by two durable and non-durable inputs.

Critical in the determination of optimal service extraction rates from durables are the concepts of user cost, part of which is the change in ending salvage value of the durable as a result of use. This cost was formalized by Neal (1942) after Keynes conceived of it in 1936. A second user cost involved in extracting services in the current time period is the discounted value of future services foregone by current use. This cost was identified by Lewis (1949).

Baquet integrated these concepts with the replacement concepts of Perrin to determine the optimal economic conditions for his model. He then derived the first-order conditions for optimality with respect to non-durable inputs, maintenance, service extraction rates, and calendar life of the durables, although he suggests that an iterative search routine is required to finally resolve all the unknowns.

The theory developed by Baquet has important implications for the supply response model presented by Johnson (1960). With fixed extraction rates and divergent acquisition and salvage prices, as assumed by Johnson, the firm's supply curve has discontinuities. The theory developed by Baquet would suggest that when variable extraction rates are considered, the discontinuities would be reduced or eliminated since firms could respond to changes in output prices by altering the intensity of use of their durable assets rather than by acquiring additional units of the durables or disposing of existing durables.

Uncertainty and Decision Making

The decision-making environment facing public utilities is an uncertain one. Exogenous forms of uncertainty include uncertainty with respect to the amount and location of deposits of various kinds of natural resources and uncertainty about the cost of extracting them. The development of new technologies provides another major source of uncertainty. Since utilities must invest in advance of anticipated needs, they are subject to uncertainties surrounding changes in future demand for an energy product--say, from electrically powered cars or from significant consumer substitutions among or away from various energy products through conservation practices. Finally, they face considerable uncertainty on the price side--the rate charged customers is, in essence, controlled by public service commissions.

Uncertainties in the economy could also have major impacts. The commercial use of energy products is directly related to the economic climate. Further, government policies will impact both on the economic situation facing utilities and on the uncertainty surrounding regulatory requirements.

Thus, the ubiquitous nature of uncertainty requires our analysis to account for its presence. First, our disciplinary models require uncertainty in order to perform the marginal types of analysis, such as measuring supply response changes to changes in expected prices, etc. A second area is decision theory. Given that action choices have been identified and their outcomes described in terms of probability distributions, how are the choices ordered from least to most preferred? This poses a particular problem, since decision makers are not likely to be unanimous in their preferences. Techniques available for ordering preferences have been reviewed by King and reported in detail in Chapter V.

The Mid-Range Energy Forecasting System (MREFS)

The discussion thus far has focused on firm-level decision models. In contrast, the Department of Energy's Mid-Range Energy Forecasting System (MREFS) is a system of models dealing with the entire energy sector of the United States. It deals with the importation of energy in addition to the discovery and exploitation of energy sources. It has components dealing with all major sources of energy, except possibly photosynthetic generation of energy from carbohydrates in agriculture. MREFS also models transportation, refining, electricity generation and the other processes between the acquisition of basic energy feedstocks and their eventual consumption by final consumers. There is also a major demand component in the MREFS which deals with the demand for various energy products in different regions of the U.S. The focus of this report, however, is on the adequacy of that part of MREFS which deals with electricity supply, demand and utilization.

The utilities component of MREFS is an aggregative system. Individual utilities are not modeled; instead, the subsector is modeled as if it were under the control of a single maximizing decision maker. This is likely an inaccurate model of the utility subsector, because the electric utilities are regulated by a wide variety of governmental units and, hence, are by nature locally monopolistic and do not act as if under the control of a single decision-making body.

The utilities model in MREFS is a single-period, non-recursive linear program which models price responses in each of ten regions of the U.S. The responses to prices are in the forms of (1) changed rates at which existing plants are operated; (2) the building of new plants; (3) the retrofitting of existing plants in response to environmental regulations; (4) the conversion of existing plants from one kind of fuel to another; and (5) the retirement of plants. MREFS responses are computed for a base year and a target year, under assumed conditions for each year, with no consideration of how the system adjusts "from here to there" over time.

On the demand side, econometrically estimated, continuous, log-linear, constant-elasticity demand functions are used for various energy products. These functions are converted to step functions for use in the MREFS integrating linear program model, which links demand and supply to arrive at equilibrium prices and quantities.

It is concluded that both the supply and demand sides of the electric utilities component of MREFS suffer from inadequate:

1. conceptualization of the dynamics of investments and disinvestments which shift demand and supply curves in response to changing economic, energy and regulatory conditions;
2. attention to changes in the rate at which energy consumers and energy suppliers extract services from fixed investments in response to changes in price associated with technical, institutional and human changes; and
3. attention to sequences of events between base periods and projection points, particularly those resulting from investment and disinvestment decisions (including mistakes) made by both the electric utilities and their customers.

Investment, Disinvestment, and Use of Durables: An Analytic Framework

The production rule followed by utilities in supplying an energy product is: utilize durable and non-durable inputs in such a manner that rates charged customers are minimized. So, how do utilities organize to meet customer demands in a least-cost manner? Obviously, capital investment/disinvestment decisions have an important impact on costs, but capital investment/disinvestment decisions cannot be made independently of decisions regarding the use of capital items and decisions regarding the purchase and use of non-durable inputs (inputs used up in a single period).

In what follows, an analytic framework is developed for prescribing capital investment/disinvestment decisions for durables. In addition, optimum conditions for replacement and use are identified. To obtain such results requires, first, a detailed examination of costs associated with extracting services from the durable.

Cost Categories

Two broad classes of assets identified in this report are durables--assets which provide services for more than one period--and non-durables--assets used up in a single period. If an asset is durable, then it has a capacity to deliver services that are held in inventory from period to period, and the costs associated with the durable result from: (1) altering the quantity of available services; (2) holding a quantity of available services in inventory; and (3) altering inventory costs in the future by current-period use decisions.

Capacity Costs

The first category of costs--defined in this report--is called capacity costs. It is very much like the cost (identified by Keynes) of using up the durable. Such a cost requires we answer two questions: (1) what is used up, and (2) how is it used up?

"What is used up" is a question that must, in some sense, be answered uniquely for each durable because each durable may provide a unique service. Passenger cars provide miles of transportation services, generators provide kilowatt-hours of electrical energy, light bulbs provide lumens of light, etc. What is used up is the capacity--a lifetime capacity--of the durable to provide services.

It is likely the case that the total amount of services extracted from the durable over its lifetime depends on the rate at which services are extracted from the durable. The potential range of service extraction rates is defined as the operating capacity of the durable. The operating capacity that minimizes average loss in lifetime capacity is defined as the durable's rated capacity. To extract services at the durable's rated capacity is to utilize the durable at a 100 percent rate. To extract services at some other rate implies a capacity utilization rate less than or greater than 100 percent.

There are at least three ways the lifetime capacity of the durable can be altered: (1) through use; (2) as a result of the passage of time; or (3) through maintenance. The relationship between losses in lifetime capacity and the rate of utilization, or service extraction rate, is determined by the design of the durable. In some cases, the average loss in lifetime capacity is only slightly affected by changes in the utilization rate--these durables we define as flexible. For those durables whose lifetime capacity is reduced by time and, possibly, whose marginal loss in lifetime capacity is not constant, average losses will be affected--and the durable is non-flexible.

Flexibility is an important issue for utilities. If the amount of services required from the durable varies, then it is more important to be able to vary the utilization rate of the durable without markedly changing the average loss. If, however, output is constant, then the concern in selecting the durable is to find one with the minimum average loss for the desired output.

Utilities which face a time-varying demand for their product, in essence, design for flexibility by investing in both "base" units and "peaking" units. Base units are durables whose rated capacity is near the upper limit of the range of operating capacities. This is because the marginal loss in lifetime capacity of extracting services is nearly constant and large time-related costs are involved. Peaking units, on the other hand, have increasing or non-constant marginal losses associated with use and smaller time costs and, as a result, have rated capacities below the upper range of their potential service extraction rate.

Maintenance performed on durables can be considered like a durable itself, if the maintenance service lasts beyond a single period. Maintenance alters the capacity of the durable, either by altering the rate lifetime capacity is lost through use or by increasing its capacity by replacing worn parts. As a result, determining optimal maintenance is a problem very much influencing the determination of the optimal investment/disinvestment patterns of durables.

Inventory Costs

Earlier, we distinguished between durables and non-durables based on whether or not they provided services beyond a single time period. Because a durable has a life beyond a single period, it has costs and benefits in common with all inventories of assets. Let's consider two.

To commit resources to an inventory is to forego the returns that could have been earned in another investment. To own a $5,000 car is to forfeit the return $5,000 could earn elsewhere (or save the cost of borrowing it). This cost we refer to as a control cost.

On the other hand, to purchase a car this year for $5,000 and to find out that next year the same car costs $5,500 is to save $500. Of course, one may not be so fortunate in his choice of investments (buying a gas guzzler just before the oil embargo) and find that the price of his durable has declined over time. Nevertheless, whatever the price change in the durable and for whatever reason--be it changes in the cost of non-durable inputs, imperfections in the market, technological changes, etc.--the cost (benefit) associated with the price change, independent of that price change which occurred as a result of use, is a time depreciation cost.

Indirect Capacity Costs

There is an important interrelation between capacity costs and inventory costs which can best be described with an example. Suppose the durable in question is a large quantity of grain stored on the owner's farm. The owner needs to decide whether to sell it, feed it, or hold it in inventory until next year. To determine the optimal disposition of the durable this period, he asks: what do I give up by feeding it or selling it this period compared to holding it in inventory? In the first case, disposing of it this period eliminates control costs in the coming period (a cost savings), but it also eliminates the impacts of future time depreciation costs, which may be favorable--i.e., an appreciation. Therefore, the impact of current-period use decisions on future-period control and time depreciation costs is an important consideration in making current-period decisions. This cost is called indirect capacity costs.

Other Costs

There are other cost considerations that may be relevant in determining optimal investment, disinvestment, and use decisions. One may be the rate at which non-durable inputs are used up. For example, the rated capacity of a durable may be 65 miles per hour, but the average loss in the amount of gasoline may occur at 45 miles per hour. Obviously, the relative costs of gasoline versus lifetime capacity of the durable ultimately determine the optimal service extraction rate.

Another cost of extracting services from the durable is the losses both in the lifetime capacity of the durable and losses in non-durable inputs associated with varying the rate at which services are extracted from the durable. An example of how such costs affect everyday decisions can again be illustrated with our car example. To travel to a destination through town is shorter--more direct--but it involves stopping at several intersections. Because of the inefficiencies and loss of fuel economy involved in stopping and starting, the freeway route to the destination is preferred even though it is longer.

Our final cost consideration we define as replacement opportunity cost. To describe this cost, we again return to our car example. Suppose our current car is rather fuel inefficient and the price of gasoline is increasing, yet the benefits still exceed the costs so we continue to drive it. Then, assume the alternative is a newer subcompact car that is quite efficient compared to the durable in use. If we account for all the costs described thus far, we will have missed the one perhaps most significant--the average difference between the costs and benefits obtainable from the replacement.

If the optimal life of the durable is 10 years, to postpone by one period the acquisition of the new durable one period is to postpone by one period the benefits of those services in each of the 10 years. It is the average benefit of those 10 years, then, that is given up by not replacing--and it is this cost we define as replacement opportunity cost.

The Indeterminacy Problem

Now, we face up to a rather vexing problem in durable investment, disinvestment, and use analysis. To determine optimal service extraction rates requires we know indirect capacity costs; but to know indirect capacity costs requires we know the optimal life of the durable; and to know the optimal life of the durable requires we know replacement opportunity cost; and to know replacement opportunity cost requires we know the optimal service extraction rates and optimal age of the replacement; etc.

To avoid the indeterminacy problem, this report proposes the following iterative approach: (1) Determine the replacement opportunity cost based on the existing durable's performance, and begin the analysis procedure by guessing the durable's optimal age. (2) Then, with the durable's life determined, calculate the optimal service extraction rates and, where relevant, the optimal values for non-durable inputs. (3) Next, identify the net returns in each period by subtracting from gross returns capacity and inventory costs. (4) Then, compare the resulting net return for the last period with the annualized average returns for the entire life of the durable (replacement oppportunity cost). If the return in the last period is less than (greater than) the annualized average, the life length selected was too long (too short) and so the procedure is repeated, beginning with a new choice for the life of the durable. This procedure is repeated until the optimal life is finally found, i.e., the one for which returns in the last period equal replacement opportunity cost.

Determining the Value of Durable Services

The problem of measuring the durable's return has not been discussed yet but constitutes a critical element of our analysis. When an asset is divisible in acquisition and in use, our marginal analysis says to compare the benefits and costs associated with the last unit in order to determine the optimal amount to use. Suppose, however, as is most often the case for durables, that the asset acquired is lumpy in acquisition but divisible in use. Then, the marginal approach is no longer relevant, except to determine the optimal units of service to extract. What is needed is a measure of the total value of services to be extracted in order to determine whether the durable should be acquired and, if acquired, for how long should it be kept.

To obtain such a total value measure, this paper introduces a new theorem--the Product Exhaustion Theorem (PET). This theorem, a generalization of what Euler's theorem does for linear homegeneous functions, partitions output among the inputs in such a way that output is just exhausted (or accounted for). The amount of output attributed to each input depends on the marginal contribution of each, measured at each stage of the production process, along the expansion path.

This theorem is particularly valuable in answering many questions not possible to formulate without its results. For example, one might ask: how will the value of services from the durable change if a different expansion path is followed as a result of increases in the price of variable inputs? These implications and others are discussed in Chapter IV in connection with the PET theorem.

The Cost of Risk

Into the previous analysis, we introduce still another consideration--the cost of risk. The cost of risk we define as that cost willingly subtracted from expected income so that the difference between expected income and the cost of risk could be received with certainty. This cost, of course, depends on several factors.

For one, it depends on the decision maker's marginal utility of income: does it increase, decrease, remain constant, increase and then decrease, etc., with respect to income? The cost of risk also depends on the uncertainty of the event. This uncertainty depends in turn on the events which give rise to it, i.e., the sources of uncertainty and the possible responses to it. If output prices are uncertain but output can be varied, then an appropriate response to uncertainty is to adjust output. However, if, as is most likely the case for utilities, output itself is uncertain because it must meet an uncertain demand, then the firm's only response may be in designing durables for flexibility.

All of these considerations are discussed in Chapter IV along with the effects on use decisions if the lifetime capacity and the loss of lifetime capacity associated with use are also random variables.

Choosing Between Probability Density Functions

When faced with investment or action choices in which the outcomes are uncertain, the choices are often ordered on the basis of the expected utility hypothesis. It combines information about decision-maker preferences along with expectations concerning the relative likelihood of alternative outcomes under each action choice being considered. The result is an index, an expected utility measure that orders action choices—the largest one being preferred.

Despite its widespread acceptance as a theoretical tool, the expected utility hypothesis has not been widely accepted as a practical tool. This lack of acceptance has been due primarily to (1) the imprecision in the measurement of decision-maker preferences; (2) problems in statistical estimation; and (3) respondents' lack of precise knowledge about their preferences.

Imprecision in the measurement of decision-maker preferences has led to the development of efficiency criteria. An efficiency criterion orders action choices into efficient and inefficient sets based on rather general preference characteristics which well-defined classes of decision makers are assumed to hold. As such, an efficiency criterion can be used to eliminate some feasible choices from consideration without detailed information about the decision maker's preference.

The weaknesses of most efficiency criteria have been that they are not adaptable to decision makers' risk attitudes. They require, instead, that decision-maker preferences match the assumption(s) underlying the efficiency criterion. For example, first-degree stochastic dominance (FSD) and second-degree stochastic dominance (SSD) both assume that the decision-maker set for whom they are ordering choices includes the decision maker who is the most risk averse of all. Thus, the efficient set must include choices preferred by the "maximin" decision maker even though, in practice, the decision-making set did not include him.

A newly developed efficiency criterion (Meyer, 1977a) appears to provide an alternative to the more limited ones currently in use. It is adaptable, that is, it can be made to match the preferences of the decision makers yet be discriminating to the degree determined by the researcher. However, because it is adaptable, it requires that preferences be measured.

To begin, Meyer's criterion classifies decision makers according to an upper and lower bound on their absolute risk aversion function. Preferences are described in terms of absolute risk aversion coefficients, because the underlying function, the utility function, is not unique. Having specified upper and lower bounds on the decision maker's absolute risk aversion function, for each pairwise comparison of probability functions, the question is asked: of all the possible decision makers included in the class defined by the upper and lower bound absolute risk aversion functions, which one is least likely to prefer distribution 1 to distribution 2? If the decision maker least likely to prefer 1 to 2 prefers 1, then everyone else will also. The actual solution to this problem is an optimal control program used in this project and described in King (1979).

Still, the question remains of how to find the appropriate upper and lower bound absolute risk aversion functions. A contribution of this project has been the development and testing of methods appropriate for interval identification.

The process begins with the theoretical result that, under certain conditions, a choice between two outcome distributions defined over a relatively narrow range of outcome levels divides absolute risk aversion space over that range into two regions. one consistent with the choice and one inconsistent with it. The level of absolute risk aversion at which the division is made depends solely on the two distributions, i.e., their properties define the two regions. The decision maker's preferences, as revealed by his ordering of the two distributions, however, determine into which of these two regions his level of absolute risk aversion is said to fall. By confronting the decision maker with a series of choices between carefully selected pairs of distributions, the region of absolute risk aversion space which is consistent with the decision maker's preferences can repeatedly be divided until a desired level of accuracy is attained. Upper and lower limits for the level of absolute risk aversion are determined at several outcome levels. These values are then used to estimate upper and lower limits for the absolute risk aversion function over the relevant range.

Now, having obtained a method for measuring risk preferences which consists of placing confidence intervals around the absolute risk aversion function, we can then adjust the width of the interval to obtain the desired research results. Suppose we must choose one unique action choice from a large set. Then, the desired width would be a line, i.e., the decision maker's unique absolute risk aversion function associated with his utility function. However, because of the imprecise measure of preferences, this could result in the actual preferred choice being eliminated from the efficient set (of one action choice).

On the other hand, suppose all that was needed was an initial screening to reduce a feasible set of action choices to a smaller, efficient set; then, the width of the absolute risk aversion interval could be increased. Now, most likely, more than one action is in the efficient set. This reduces the likelihod that the preferred choice would be rejected (excluded from the efficient set), while increasing the probability that a number of choices not actually preferred are in the efficient set. These types of trade-offs now become possible with the Meyer criterion.

The interval approach to the measurement of decision-maker preferences and Meyer's efficiency criterion facilitate the application of decision theory based on the expected utility hypothesis in the analysis of practical decision problems. They permit recognition of the fact that preferences cannot be measured exactly, and they allow explicit consideration of the trade-offs between the accuracy and discriminatory power associated with differences in the precision with which preferences are measured.

A Practical Model

The simplifying assumptions necessary for analytical development of the theoretical model reduce, somewhat, the usefulness of the theory in practical decision-making contexts. By applying simulation techniques and exploiting the capabilities of large-scale computers, however, many of those assumptions can be relaxed and the model made much more complex in terms of incorporating with the theory more of the features of the reality faced by decision makers.

A computer simulation model has been developed, therefore, which incorporates the various time, control, user, and replacement opportunity cost concepts as a decision aid for investment, disinvestment, and use decisions of energy supply firms faced with various types of uncertainty. In addition, practical features of the real world are included, such as, in the case of electric utility decision makers faced with capacity expansion decisions, (1) the use of multiple durables, including the distinction between base and peaking units and between units using different fuels; (2) the requirement to meet a time-varying demand, plus a reserve capacity; (3) a customer pricing formula based on fuel and other variable costs and an allowed rate of return on capital; and (4) construction and licensing lead times. Aspects of uncertainty which may be considered include, for example, demand levels, fuel prices, forced outages, and construction costs. Regulatory uncertainties are not included in the present version of the model, but nothing in principle precludes their consideration at a later date.

Recognizing the intertemporal dynamics of the investment and disinvestment in and use of durable assets, as developed in the theory, the simulation model is conceptualized as a state-space optimal control problem. No single optimal control solution is actually sought, however. Instead, in keeping with the utility theory of choice under uncertainty described above, risk aversion measurement tools are used with the criterion of stochastic dominance with respect to a function to identify an efficient set of solution strategies. For this purpose, a strategy is defined as an assumed set of decision rules and decision actions which operate over the planning horizon.

When placed in the perspective of the range of decision aids commonly used by electric utilities and other energy suppliers, the simulation model developed here can be considered to be of intermediate complexity with respect to manpower and computer resources required to use it. That is, it is more complex than such methods as the minimum revenue requirements discipline in that it can take a system-wide perspective and it incorporates the cost and uncertainty concepts of the theory described above. On the other hand, it need not include all the intricate engineering and financial detail of system or corporate planning models. Therefore, it can serve as a complement to these models to evaluate and screen decision options in the early stages of planning.

A Comparative Empirical Test

There are many ways to test a theory or model to establish its credibility. Such tests fall into four categories: (1) tests of coherence to establish internal logical consistency; (2) tests of correspondence to establish fidelity, in relevant respects, to the real world; (3) tests of clarity to establish that the model is unambiguous and comprehensible to its users; and (4) tests of workability to establish, first, cost effectiveness in use and, second, that the model or theory contributes to better decisions. Tests in all four categories were conducted on the simulation model and, thereby, on the advanced theory it contains, and a great deal more such testing is necessary--indeed, model tests are never complete in that every use is a test. We report here on one test carried out in accordance with Task 4 of the project.

Comparing the application of a new method and its results with an established one falls under the headings of both correspondence and workability testing. Interpreting the model and its results also serves as a test of clarity. For this purpose, then, the theory and simulation model were compared with the minimum revenue requirements (MRR) method commonly used in utility planning.

As a specific case, the comparison was made in the context of a decision faced in 1977 by the Consumers Power Company of Jackson, Michigan. The decision was whether to reacquire and reactivate two previously decommissioned dam sites and hydro stations on the Boardman River near Traverse City, Michigan. Using the same data as the MRR analysis, a similar conclusion was reached: the investment would yield only a marginally positive expected gain with a relatively high variance and a long payback period.

Computationally and in the treatment of uncertainty, the two approaches are similar. Both look at the discounted present value of economic gain; and the MRR method often uses Monte Carlo analysis to capture uncertainties and, further, nothing in principle would prevent its use of the approach of identifying a risk-efficient set of options, as is done with the simulation model. The major differences are in the new theory's more complete consideration of economic costs; its criterion for determining the economic lifetime of the durable, i.e., the time to disinvest or replace it; and its determination of optimal service extraction rates.

As we gained experience with the model and the Boardman application, however, we realized that, while suitable for such partial or project analyses as this, the limits of the simulation model and its embodied theory can better be explored as a component of a corporate planning model or in comparison with a capacity planning model--a more strategic level to which the MRR approach is not well suited. Such further tests are recommended.

Recommendations for Future Work

In building analytic models to develop a theory, many simplifying assumptions must be made. This project was no exception. While the simulation model was able to relax some of these assumptions and introduce greater practical realism into the analysis, further theoretical developments are nevertheless necessary to provide a solid foundation for practical analysis with improved simulation models. Therefore, the following sampling of areas requiring further theoretical advances and, thereby, simulation model extensions is discussed in Chapter VIII:

1. Conglomerate durable analysis -- to recognize that a durable, e.g., an electric power plant or a car, is frequently a system of component durables which may have lifetime and other economic decisions associated with them as well.
2. Durables in parallel and in series -- to recognize that decisions regarding one durable cannot be taken in isolation from decisions regarding other durables with which it is operating in parallel (e.g., one generating plant in a grid) or in series (e.g., generators, transformers, and transmission lines).
3. Regulations, taxes, and inflation -- to incorporate them and their uncertainty into the analysis, e.g., whether allowed rates of return should be tied to book value or replacement value.
4. Uncertainty -- to consider additional important types of uncertainty, such as intermittent supply sources, which buy-back requirements are making of increasing concern.
5. Other extensions of economic theory -- such as decisions relating to non-durables which are not infinitely divisible in use or acquisition.
6. Optimal control analysis -- to supplement the risk-efficient set approach as a check on the realism of the problem specification and results.

CHAPTER I

INTRODUCTION

Lindon J. Robison

Background

Those industries which supply energy in various forms face an uncertain environment which complicates their decision-making process. They rely on natural resource inputs whose availability and location are uncertain, making the optimal rates of extraction, use, and exploration of these resources uncertain. Moreover, the energy supply industry, while relying on natural resources for inputs, may transform these inputs into alternative energy forms. Yet, the cost and the technology available to do this are continually changing, adding still another dimension of uncertainty. In addition, the energy supply industry makes long-term investments in plant and equipment which, to a large extent, determine its capacity to meet the energy requirements of its customers in future time periods. Further, the amount of energy the industry will be called on to supply is still another random variable. In any one year, weather (a random variable) may alter dramatically the need for energy at particular times of the year. To exacerbate the problem, government regulations require that the energy supply industry be prepared to meet peak load demands, which are highly price inelastic, whatever they may be.

Because of the ubiquitousness of uncertainty in the energy sector, models which ignore it can be expected to inadequately explain supply response to price changes, unless uncertainty has no impact on decision makers. The evidence suggests, however, that uncertainty has important impacts on investment and supply responses of decision makers, particularly in the energy supply industry. For example, uncertainty about the amounts and locations of natural resources will likely result in slower extraction rates to conserve supplies until other resource supplies can be confirmed. On the other hand, uncertain peak load demands may result in over-expansion or under-expansion of energy-producing plants, depending on the relative costs of not meeting customers' peak load requirements versus the opportunity cost associated with under-utilized plants and equipment. Or, finally, uncertainty about what changes may occur in environmental standards that regulate energy-producing plants and equipment may affect new plant and equipment investments needed to produce sufficient energy supplies in the future. Possible changes in standards that require costly modification to existing durable stocks, or that render them obsolete, are still another important source of uncertainty that determines the supply response of the energy industry.

Uncertainty is also related to asset fixity. An asset is fixed, that is, its holdings are not varied by the firm, as long as its marginal value product is bounded above and below by its acquisition and salvage values, respectively. Under these conditions, firms have no incentive to invest or disinvest in those assets meeting the marginal value product bounds. However, what if the marginal value product generated by the durable asset, maintenance requirements, and its acquisition and salvage values are random variables, affected by all the sources of uncertainty described earlier? Does the durable become fixed over a wider range of production? Or does it reduce asset fixity? Obviously, the answer has important impacts on the outcome of any study designed to measure energy supply responses to changed incentives and, therefore, needs to be explored.

In order to examine the combined effects of uncertainty and asset fixity on the supply of an energy product, the Electric Power Research Institute (EPRI) funded RP 1220-2, the results of which are described in this report. The primary objective of 1220-2 was to improve energy supply forecasting models by integrating the theory of decision making under uncertainty with the theory of investment, disinvestment, and use of durable assets under uncertainty. A model providing such integration of theory can more accurately describe the ability and willingness of the energy supply industry to respond to changes in the probability distribution for different energy output and input prices.

Research Objectives

Specifically, this project had as its objectives:

(1) To improve energy supply forecasting models by integrating the theory of decision making under uncertainty with the theory of investment and asset fixity;

(2) To demonstrate how the integration can more accurately describe how energy supply industries respond to changes in prices and their probability distribution; and

(3) To provide modifications and methodological improvements for the incorporation of uncertainty and asset fixity in energy supply models.

Research Tasks and Report Overview

To achieve the objectives of this project, the following tasks were carried out. Task 1, reported in Chapters II and III, reviewed selected previous studies on the causes of supply variations in the major energy-producing industries to obtain, where possible, empirical measures that would aid in determining the relative importance of variables producing energy supply variations. This review includes literature pertaining to:

(a) the effects of regulations on the pricing system for energy services;

(b) the causes of supply variations in the supply of an energy product;

(c) decision-making procedures used within the energy industry; and

(d) the importance of asset fixity and uncertainty on the supply of a particular energy product.

In addition to reviewing literature on the empirical importance of supply variations, selected theoretical models are reviewed for their usefulness in explaining the observed variations in energy supplies. These models include:

(a) the effects of asset fixity on supply response;

(b) the effects of uncertainty on asset fixity;

(c) theoretical models which include risk components; and

(d) theoretical and empirical models that describe the decision process used within the energy supply industry.

The review of the Department of Energy's Mid-Range Energy Forecasting System (MREFS) in Chapter III supplements the reviews of the micro models described in Chapter II. The review points out key relationships between micro and macro models used in energy supply analysis.

The review in Task 1 points out the need to extend the existing theory of decision making under uncertainty. This extension involves combining the theory of investment, disinvestment, and use with the theory of decision making under uncertainty. This leads to Task 2.

Task 2, reported in Chapters IV and V, was to extend the existing theory of decision making under uncertainty to more accurately describe the processes whereby an energy supplier acquires inputs and makes investment and disinvestment decisions, which ultimately determines its ability to respond to energy demands. In meeting the requirement of Task 2, this project recognized two different areas in the analysis of uncertainty: namely, the application of uncertainty to economic models and the development of criteria for ordering action choices described by probability density functions. In Chapter IV, the analysis begins with a very careful theoretical development of the concepts of cost and benefits of using and owning durable investments. This distinction, as well as the application of user cost concepts to the extraction of services from a durable, the identification of capacity, and a formal statement of cost, constitute a unique contribution to economic theory made in this report. Going beyond earlier work in durable analysis, Chapter IV also proves a theorem that allows one to value durable services, determine the optimal rate of service extraction, and determine the optimal lifetime, that is, the time to disinvest in each durable. In essence, it recognizes that the concept of asset fixity is a time dimension problem, and that determining when an asset is to be disinvested is essentially answering the question: <u>How fixed is an asset?</u>

Uncertainty is introduced into the analysis by treating demand, the cost of using durable services, and the availability of services for the durable as uncertain. Economic consequences are then deduced.

Chapter V reports the development of a new theoretical tool for ordering uncertain prospects. As such, its usefulness is in empirical models which, in effect, describe in probabilistic terms the outcomes resulting from alternative action choices. In describing the usefulness of the new decision theory tool--stochastic dominance with respect to a function--Chapter V draws an analogy between Type I and Type II errors. Type I errors, the rejection of a preferred action choice from the efficient set, are more likely to occur with single-valued utility functions. Type II errors, the inclusion of unpreferred action choices in the efficient set, are more likely to occur with the efficiency criteria defined for arbitrary classes of decision makers. Chapter V presents a flexible alternative to these two extremes which proves to be a valuable tool in the modeling effort described in Chapters VI and VII.

The third task of this report required an empirical model be developed to forecast energy supplies for particular energy products. Chapter VI incorporates the theoretical developments in Chapter IV and combines them with the decision tool introduced in Chapter V to provide an analytic framework for analyzing investment-disinvestment use decisions. This model is of intermediate complexity and is not intended to replace larger models available to most utilities. Nevertheless, it can be used as a preliminary planning tool. A User's Manual documenting the model has been prepared and submitted to EPRI as a separate report.

One way to evaluate new modeling approaches, especially ones intended to be useful in practical decision-making settings, is to compare its content and performance with an existing model. Chapter VII reports such a comparison, which is the requirement of the last task, Task 4. The comparison involves an investment decision which faced the Consumers Power Company of Michigan in 1977: whether to reacquire dam sites and reactivate hydro stations along the Boardman River near Traverse City. Their evaluation of this project was based on the widely used and familiar minimum revenue requirements method. To test the theoretical developments and usefulness of the models developed in this report, we employed the newly developed model using Consumers Power data. The results of those comparisons are presented in Chapter VII. The most striking conclusion was the similarity of the results, suggesting that, while not compatible in all respects, the minimum revenue requirements method and the investment-disinvestment criteria developed in this paper are compatible.

This report concludes in Chapter VIII with recommendations for future work. Obviously much remains to be done. The directions which may be followed in developing the theory of investment-disinvestment and use as well as decision making under uncertainty are detailed in that chapter. Where possible, applications to investment-disinvestment decisions facing utilities are emphasized.

CHAPTER II

EFFECTS OF UNCERTAINTY ON ENERGY SUPPLY RESPONSE

Alan E. Baquet

Introduction

Those industries which supply energy in various forms have characteristics which make them somewhat unique. This project was developed because of an interest in analyzing two of those characteristics, namely, the uncertainty which surrounds the decision-making process in energy supply industries and the capital-intensive nature of those industries. Both of these characteristics have important implications for the ability of industries to alter their supply of energy in response to changes in economic and/or political conditions. The primary objective of this study is to improve analysts' ability to describe how energy supply industries respond to changes in their environments.

This paper represents a report on Task 1 of RP 1220-2. Task 1 involves reviewing selected previous studies on the causes of supply variations in the major energy-producing industries in an effort to determine the relative importance of variables producing energy supply variations. Task 1 also involves the review and evaluation of selected theoretical and empirical models to determine their usefulness in explaining the observed variations in energy supplies.

Characterization of the Industry

As indicated above, there are several characteristics of energy supply industries which make them somewhat unique. While the focus of this study is on the uncertainty and the capital intensity of the industry, it is important to identify other characteristics that are also related to the industry's ability to alter supplies of energy.

In a perfectly competitive industry, the forces of supply and demand interact in the marketplace to jointly determine equilibrium price. The electric power industry differs from this ideal in two significant ways. First, price is not determined in an equilibrium context. Rather, the structure of rates which can be charged by suppliers of electric power is determined in a political arena. There is ample economic literature on the regulated firm which provides a theoretical basis for modeling electric utilities. This literature is discussed in following sections of this paper.

In addition to having rates determined in a political arena, there are other political decisions which affect electric power suppliers. There are fairly stringent regulations governing the environmental impacts of generating electricity. The uncertainty surrounding the establishment of environmental regulations is a major concern to the suppliers of electric power (Boris, 1977; Keady, 1977) because environmental regulations affect both the cost and feasibility of electric power generation processes.

The second difference concerns the interaction between supply and demand. In competitive industries, the price mechanism serves as the signal between supply and demand. The prevailing view in the electric power generation industry is that suppliers must be able to meet the quantity demanded. As such, the expected demand for electricity plays an important role in determining the supply of electricity. Suppliers of electricity face a continual and time-varying demand for their product. The demand for electric power varies by time of day as well as seasonally within a given year, and suppliers are required to adjust their production processes so as to meet both the daily and seasonal peaks in demand. Previous work has been done on the problems imposed by this "peaking problem."*

An important implication of this demand-supply interaction involves the capital investment decisions made by the suppliers of electric power. The firm's ability to meet peaks in current demand is a result of capital investment decisions made in previous years. By the same token, ability to meet future peaks in demand is determined in part by current investment decisions. This future demand is an important component of current investment decisions.

Capital Nature of the Electric Power Industry

Because of the capital intensity of the electric power supply industry, decisions regarding the acquisition, use, and disposal of durable assets are of major importance for managers of electric power generating firms. Managers are faced with two interrelated decisions concerning capital assets (durable assets). They must decide about the optimal amount of services to extract in each production period and the optimal stock of durable assets. For an existing firm, the optimal stock may involve additions to the stock of durable assets and/or decreases in the initial stock of durable assets. Even though the acquisition/disposal decisions relative to durable assets are not made independently of the usage rate decisions, the two decisions are separated here for discussion purposes only.

---

*The Bell Journal of Economics has several articles concerning the peak load problem and the consequences for pricing.

Investment/Disinvestment Decisions

Decisions concerning the acquisition and/or disposal of durable assets are inherently different from decisions regarding the acquisition of non-durable assets. Durable assets are typically available in discrete units. Furthermore, durable assets are capable of being used during future production periods. Thus, decisions regarding the acquisition/disposal of durable assets require information about the future production periods. This is not the case for non-durable assets, since, by definition, they are entirely used up in one production period. Thus, the decisions regarding their acquisition typically do not require information about future periods.

Much of the economic literature on the theory of investment and disinvestment decisions is reviewed in Baquet (1977a). As discussed therein, the theory of investment decision making has its origins in Alfred Marshall's Principles of Economics. The basic investment decision rule advanced by Marshall still holds: Balance current costs against discounted future gratifications. Later writers, notably Frank Knight, refined Marshall's decision rule, but the basic rule still holds. A review of the engineering economics literature indicates that the basic rule stated by Marshall is used in practice (Smith, Gerald, 1968).

Practical application of Marshall's basic decision rule in the public utility industries has relied on what is referred to as the minimum revenue requirements discipline (MRRD). Paul H. Jeynes is a leading proponent of this method of evaluating capital investment decisions.

According to Jeynes, the primary objective for the firm is to discover the alternative investment plans which will result in the minimized outlays throughout their service life. He identifies two types of outlays: (1) the initial capital investment; and (2) periodic expenses thereafter, such as taxes, operation and maintenance expense, administration expense, etc.

Minimum revenue requirements are defined to be the revenues which must be obtained in order to cover all expenses incurred, associated with and including the company's minimum acceptable return (MAR) on investors' capital (Jeynes, 1968, p. 62).

Important points to note are:

(1) Only a special portion of revenues is included. These revenues must equal certain exactly defined outlays. The recovery of these outlays in the form of revenues will enable the company to break even.

(2) Actual revenues are not estimated. Actual revenues are expected to be greater than the minimum revenue requirements.

The following table from Jeynes (1968, p. 61) is reproduced to illustrate these points. Note that C + D + E is the same as 2 + 3 + 4, thus A + B is the portion of revenue needed to recover 1, the initial capital investment less net salvage.

Table 2.1 Outlays Over the Service Lifetime of a Project
(Which Are To Be Minimized)

Expression I. Outlays (= Disposition of Funds)

1. The initial capital investment (= purchase price installed)
   a. Less ultimate net salvage, when received
2. Taxes
3. Operation and maintenance expense
4. Other expenses, such as administrative and general expense and sales and collection expense

Expression II. Revenue Requirements (= Source of Funds)

A. Minimum acceptable return on the capital investment (not including any profit)
B. Retirement cost (or amortization, commonly known as "depreciation expense," adjusted for ultimate net salvage)
C. Taxes
D. Operation and maintenance expense
E. Other expenses, such as administrative and general expense and sales and collection expense

Both of these expressions describe the same minimum revenue requirements of a project.

Profit incentive is defined to be the difference between actual earnings and the minimum revenue requirement. The components of minimum revenue requirements are: (1) minimum acceptable return on capital investment, (2) depreciation, (3) taxes on MAR, and (4) operation and maintenance expense, and other periodic expenses. The determination of these components will vary by project; however, guidelines can be specified as to the intent for each component.

Minimum Acceptable Return

If a project is to be undertaken, its earnings must be greater than the cost of financing which is to be minimized. It is the long-term minimum acceptable return on the company's pool of investors' capital that is in question, not the minimized cost of immediate financing of the new project. The fundamental objective is to determine the company's MAR as a percentage of investors' committed capital. For external financing, this can be expressed as:

$$\frac{d + X\% \text{ of } P}{P}$$

where d = current dividend rate

P = market price per share

X% = current rate of annual increase in market price

Jeynes suggests that these "instantaneous" estimates of MAR be averaged over a relevant time period in determining the long-run MAR.

In the paper by Boris (1978) presented at the EPRI workshop on capital investment decisions in May, 1978, it was pointed out that the weighted cost of capital was used by the Consumers Power Company in determning their MAR.

Depreciation

While MAR represents the return on the capital investment, the depreciation annuity represents the return of the capital investment. In practice, depreciation, in percent, is derived from a table of values based on (1) percentage MAR, (2) probable service life in years, (3) the retirement-dispersion pattern, and (4) ultimate net salvage.

For purposes of investment decisions, the depreciated "value" at any date short of total life is not important. The lifetime-levelized revenue requirement is the important aspect. Thus, the depreciation charge developed here will only correspond to the "book charges" for depreciation by accident.

The percentage MAR was discussed above. Estimates of probable service life, in years, is based on analyses of past and current experiences in addition to appraisals of the future. Retirement-dispersion patterns have been developed through engineering studies. Estimates of the ultimate net salvage can be based on experience and expectations about the future. With these pieces of information, the revenue required for depreciation can be estimated. This revenue requirement is simply the annualized equivalent of the initial investment less any net salvage.

Taxes on MAR

For purposes of economic analysis, the MRRD specifies two components for taxes: (1) the revenue requirement for income tax, which is the tax on the taxable portion of MAR; and (2) the income tax on the profit incentive, which is the tax on earnings in excess of MAR.

The first component is the one of interest for economic comparisons according to the MRRD. The revenue requirements for taxes will vary by project and by state. Inasmuch as taxes are law, the revenue requirements can be developed from the appropriate tax statutes.

Periodic Expenses

In addition to covering the above expenses, revenue is also required to cover the periodic expenses associated with operation, maintenance, and other expenses. The magnitude of these expenses will vary by project. Estimates for these expenses over the service life of the project are required.

Weaknesses of the Minimum Revenue Requirements Discipline

After the four components of the minimum revenue requirements are developed, it is possible to specify the minimum revenue that is required from the project to cover outlays for the project. To the extent that the actual revenues from the project exceed the revenue required to cover outlays, a profit will be earned on the project. Those projects with a positive profit should be undertaken by the firm.

The MRRD method provides a framework for making capital investment decisions. However, within that framework, some important aspects of investment decision making are ignored. For example, the impact of uncertainty is not addressed explicitly. Any investment decision which relies on information about future periods is faced with uncertainty. Of the four components in the MRRD approach, MAR is probably known with the greatest certainty, and even its determination can change over time. The depreciation requirement depends on an uncertain service life, an uncertain retirement-dispersion pattern, and an uncertain ultimate net salvage. Even the calculation of taxes on MAR may change over the life of the project.

It is this author's opinion that the most severe problem with the MRRD method is its treatment of periodic expenses. Operation and maintenance expenses may be highly variable over the service life of the investment project. Even if the variation in these expenses is known with certainty at the time of the investment decision, it can alter the relative merit of alternative investment projects. Although nothing in the MRRD method seems to exclude variable periodic costs, not making explicit reference to the possibility of varying costs is unsettling.

Recent theoretical work by Baquet (1978) does not rely on the fixed usage and, hence, cost assumption, and therefore could form a basis for modifying the MRRD method. Permitting the durable to be used at different rates over its productive life, while being a more accurate representation of reality, complicates the calculation of the durable's value in use, since the value in use is now more closely tied to future production decisions. Furthermore, when the durable asset's usage is allowed to vary and maintenance is considered, the productive life of the asset is no longer fixed but also becomes a variable.

Investment decisions within the electric power generating industry are affected by conditions in the general economy. In particular, the rate of inflation has important implications for the cost of producing electricity in future periods. Also the cost of borrowed capital affects the weighted cost of capital, the MAR, and hence the discount factor.

As indicated above, the four pieces of information which are needed in the investment decision are seldom known with certainty at the time that an investment decision is made. The specific types of uncertainty that exist are discussed below.

Production Decisions

Investment/disinvestment decisions are related to the decisions made in each production period. This section discusses the types of production decisions faced by managers of electric power generating firms. Important aspects of these decisions are (1) the supply-demand interaction, which involves adjusting supplies to meet peak demand; and (2) the usage of durable assets.

Within each production period, the electric power generating firm is expected to supply the quantity that is demanded. When demand is variable, meeting it involves adjusting the usage rate for durable assets as well as the quantities of non-durables used.

There have been studies which have compared the costs of operating at full capacity (meeting peak demand) with the costs of operating at less than full capacity (off-peak periods). The primary focus of these articles has been on deriving optimal pricing schemes rather than the effects of altering the usage rates for the durables on the investment/disinvestment decisions (Joskow, 1976; Wenders, 1976).

It is not clear that these studies considered all the relevant costs. For example, the user cost of capital is oftentimes not included. As conceived by Keynes and developed by Neal, Lewis, and Baquet, user cost is associated with the opportunity cost of using the durable asset in current rather than future production periods. The inclusion of this aspect of the cost of generating electricity in the current period may have important implications for the pricing mechanisms used by regulatory commissions.

Suppliers of electricity have attempted to smooth the demand that they face by offering incentives to large users for maintaining a reasonably constant use. These load negotiations have primarily focused on industrial users. However, some attempts have been made to offer sufficient incentives to residential consumers so that they will alter their usage patterns also. The implication of these load negotiations for the usage rate of durable assets has not been fully explored.

As is apparent from the above discussion, investment/disinvestment decisions and production decisions are interrelated. This interrelationship is particularly important for the electric power generating industry, where the usage rate for the durable assets is highly variable. In essence, the investment decisions place an upper bound on future generating capacity.

The nature of durable assets in electric power generation is such that, once they are in place, they tend to remain fixed for long periods of time. Thus, production decisions are primarily concerned with how best to use the fixed capacity to meet demand.

The Effects of Asset Fixity on Supply Response Models

Economic theorists have long recognized the limits that fixed factors of production place on a firm's ability to respond to changes in its economic environment. Early theoreticians, however, considered assets to be fixed in a technical or physical sense rather than an economic one. It was not until the early 1950s, when Glenn L. Johnson, Joseph Willet, Lowell Hardin, and others recognized that an asset can and generally does have different acquisition and salvage values, that an economic definition of asset fixity was developed. Edwards wrote, "an asset is fixed if it isn't worth varying" (Edwards, 1958, p. 15). In economic jargon, this statement indicates that an asset is fixed if its marginal value in use is bounded above by its acquisition price and below by its salvage price. Edwards (1959) explored the consequences of this definition.

Johnson (1960) saw the supply response implications of acquisition/salvage price differentials. The salient points of Johnson's article are presented below.

Consider a production function of the form $Y = F(X_1 | X_3)$. Johnson considered X_3 to be a durable asset for which the following condition holds:

$$(2.1) \quad Px_3a > \frac{\partial Y}{\partial X_3} P_y > Px_3s$$

where Px_3a = acquisition price for X_3

Px_3s = salvage price for X_3

Py = price of output Y

$\frac{\partial Y}{\partial X_3}$ = marginal physical product of X_3 in producing Y.

The conditions imposed by (2.1) imply that X_3 is fixed for the entire range of possibilities to be considered.

Now, for an input X_1, we have $Px_1a > Px_1s$, so the amount of X_1 on hand, depicted by a in Figure 2.1, is fixed under some conditions but variable under others.

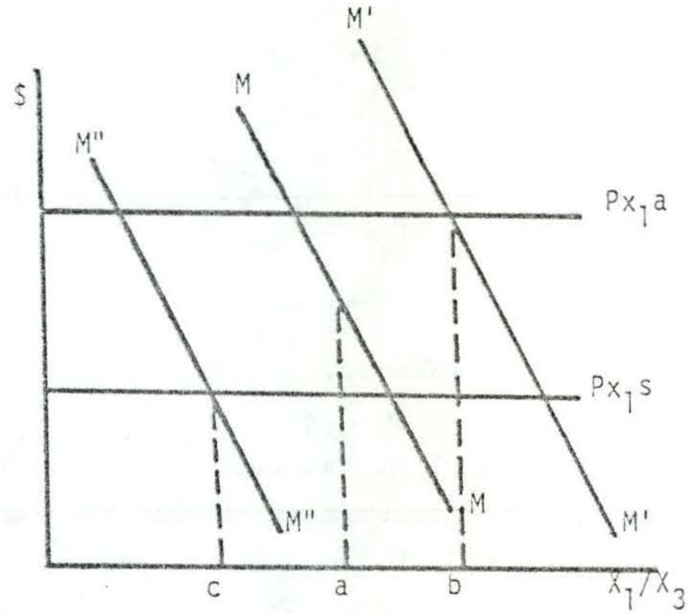


Figure 2.1

Acquisition and Salvage Prices and the Definition of Asset Fixity

The marginal value product (MVP) of X_1 is given as $\frac{\partial Y}{\partial X_1}$ Py, which is MM initially. Thus, X_1 is fixed at a since $Px_1a > MM > Px_1s$. If, however, Py increases so that $\frac{\partial Y}{\partial X_1}$ Py shifts to M'M', the MVP of X_1 exceeds its acquisition price and the most profitable amount of X_1 to use increases to b. A fall in Py which drops the MVP of X_1 to M"M" would reduce the most profitable level of X_1 to c. As Johnson states it, the definition of asset fixity for X_1 is that X_1 is fixed whenever $Px_1a \geq \frac{\partial Y}{\partial X_1} Py \geq Px_1s$.

The divergence between acquisition and salvage prices for X_1 gives rise to a discontinuity in the marginal cost of producing Y, MCy, at the point $Y = F(a \mid X_3)$. Beyond this point, $MCy = Px_1a / \frac{\partial Y}{\partial X_1}$. For quantities of $Y < F(a \mid X_3)$, MCy is lower than for $Y = F(a \mid X_3)$. Figure 2.2 depicts the marginal cost curve.

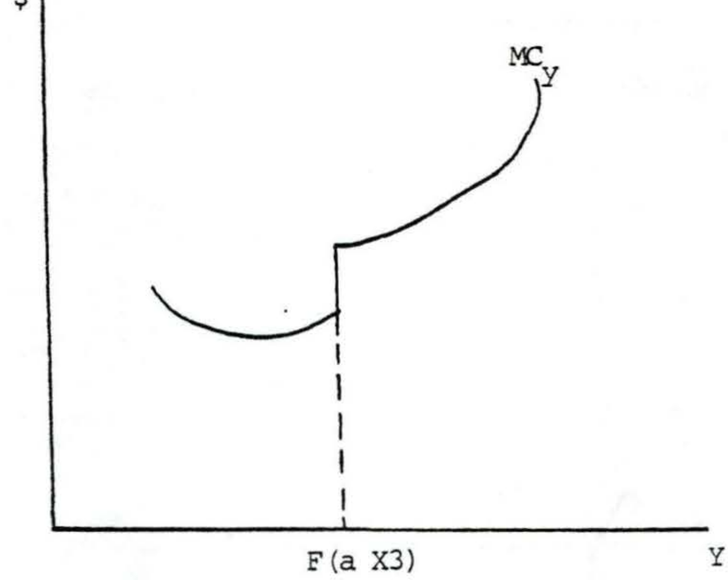


Figure 2.2

Asset Fixity and a Discontinuous Marginal Cost Curve

Johnson extends the analysis to the two variable input cases. He derives lines of least cost combination for expansion when Py is increasing and lines of least cost combination for contraction when Py is decreasing. As Johnson indicates, the marginal cost curve associated with these movements in Py has the following shape.

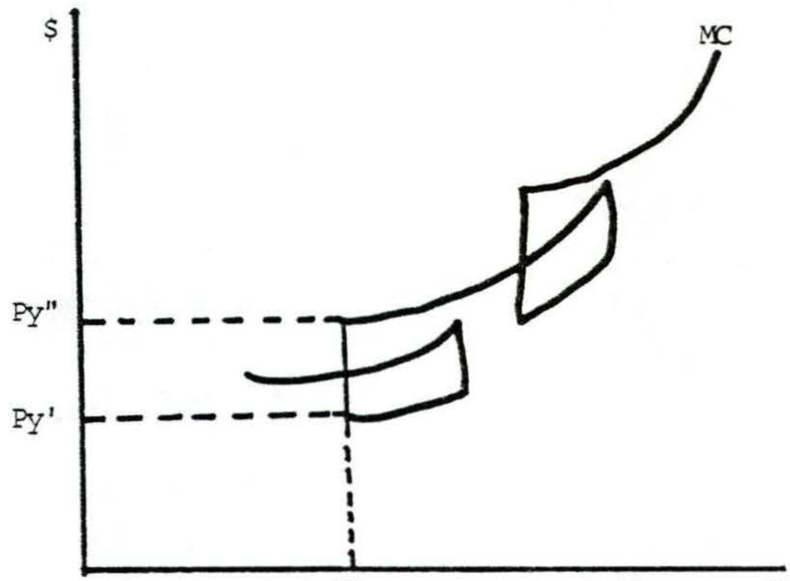


Figure 2.3

The Marginal Cost Curve and Changing Product Price

The vertical portions of the curve in Figure 2.3 indicate that both variable inputs are fixed in the economic sense and that changes in output price over these ranges will not induce a response in supply. For example, if the firm is currently at Y' and price is at Py', the price must increase to Py" before the firm will increase production.

While Johnson's article laid the theoretical groundwork for incorporating asset fixity in supply response work, there are some aspects of durable assets that he recognized but did not develop. One of Johnson's simplifying assumptions dealt with the stock-flow conversion process for durable assets. Johnson assumed that there was a one-to-one correspondence between the durable stock and the flow of services from the durable. In other words, he assumed that services are extracted from the durable at a fixed rate.

Francis S. Idachaba, in an unpublished manuscript, made an initial thrust in the area of variable extraction rates. He treated the extraction rate for services as an explicit variable. It is not clear from Idachaba's writings that he fully recognized all the implications of the variable extraction rate, nor did he develop the supply response implications.

More recently, Baquet (1977, 1978) has presented a more complete analysis of asset fixity with variable extraction rates. The theoretic model used by Baquet, as depicted in Figure 2.4, considered a vertically integrated production process.

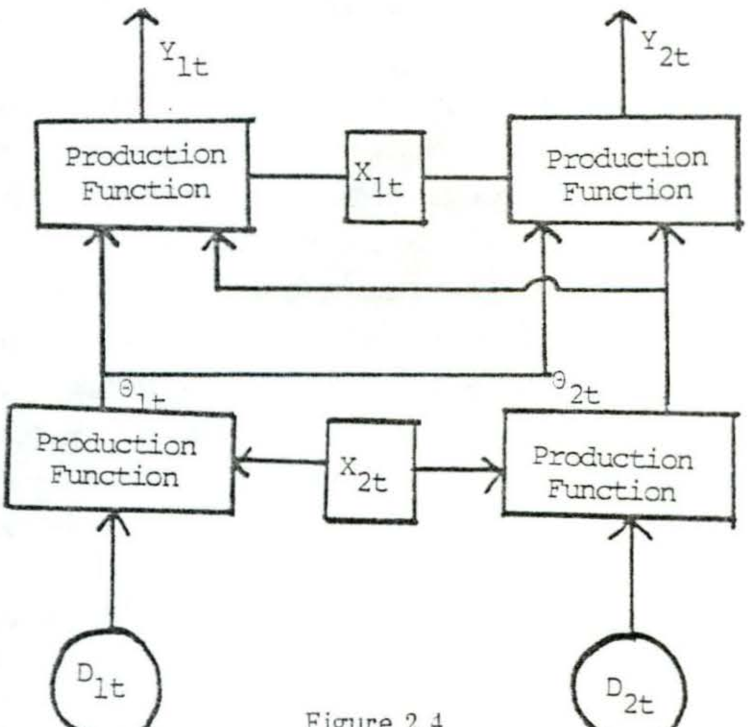


Figure 2.4

Asset Fixity and Variable Extraction Rates in a Production Process

The production process considers two durable assets, D_{1t} and D_{2t}, which are used with one non-durable input to generate services. The services thus generated are used along with a second non-durable input in the production of the final outputs, Y_{1t} and Y_{2t}. Considering the production process in this manner permits Baquet to simultaneously consider the questions of investment/disinvestment and the rate of use for durable assets.

When the extraction rate for services from durables is variable, the life of the durable is also variable. Furthermore, the role of maintenance has added significance, since it can extend the life of the durable. This relationship between services generated, maintenance performed, and the life of the durable is accounted for in Baquet's analysis.

It is assumed that the theoretical firm would maximize an objective function subject to the physical constraints imposed by the production process. The objective function used by Baquet considered the gain in the net present value of the firm in each time period as the sum of the net receipts from the current production activities plus the net gain in the value of the durable assets. The net receipts from current production activities include a component in the cost of current production which is generally overlooked. This cost is the user cost of the durable asset. One aspect of user cost is the use depreciation of the durable asset. The use depreciation may be partially or totally offset by the maintenance activities of the firm. The user cost in the objective function represents the change in the ending salvage value of the durable as a result of using the durable during the period. This cost was formalized by Neal (1942) after Keynes conceived of it in 1936.

There is a second user cost involved in extracting services in the current time period. This cost is the highest discounted value of the future services foregone by current use of the durable. This cost component was identified by Lewis (1949) and forms a portion of the disinvestment criterion for the firm considered below.

The gain in the net present value of the firm is achievable by investing in additional units of the durable, disinvesting in some or all of the units currently held, or reorganizing the usage pattern for the currently held durables. When investing in durables, the gain is the difference between the current acquisition price and the net present value of the future services generated from the durable. When disinvesting in durables, the gain is the difference between the current salvage price and the net present value of the future services generated from the durable. When reorganizing the usage pattern of the durables, the gain is the change in the net present value of the future services generated from the durables.

The optimization of the objective function involves the simultaneous determination of the optimal production activities and the optimal investment/disinvestment activities.

The optimality conditions for production activities are:

(1) For the non-durable inputs, the marginal cost of the input must equal its marginal value in use. The marginal value in use of the non-durable in the final production process is its marginal value product, while, under competitive input markets, its marginal cost is its price. Equating these two determines the optimal quantity. The marginal value in use of the non-durable used in the generation of services is determined as an instrumental marginal value product. An instrumental marginal value product is measured as the marginal physical product of the input used in the production of an output times the marginal physical product of that output used as an input in a higher level production process times the price of the higher level output. This corresponds to the concept of an input's value in a vertically integrated production process. The marginal cost of using the non-durable involves three components. The first component is the price of the non-durable. The second component is the instrumental marginal user cost of generating services. The third component is the opportunity cost of economically altering the physical life. This component is essentially an instrumental maintenance cost. The optimal quantity of the non-durable is determined by equating its instrumental marginal value product with the sum of the three marginal cost components.

(2) The optimal quantity of maintenance to perform is determined by equating the cost of a maintenance input with its marginal value product. The marginal value product of maintenance is given as the marginal physical product times the marginal value of maintenance. The value of maintenance is the value of the services which can be generated from the durable as a result of performing maintenance.

(3) The optimal quantity of services to generate in each production period is determined by equating the marginal value product of the services in producing the final outputs with the marginal cost of generating those services. The marginal cost is composed of three elements. The first element is the marginal user cost incurred by using the durable to generate services. The second element is the net cost of using the non-durable input to generate services. This component recognizes the aggregative nature of the non-durable input by taking the price of the input net of the opportunity cost of adjusting the life of the durable. The third element reflects the opportunity cost of economically reducing the physical life of the durable by generating current services, i.e., maintenance costs. The marginal cost of generating services is the sum of all three elements.

The formal mathematical expressions for the optimality conditions, which must hold at each point in time for the production activities, are given in an EPRI workshop by Alan Baquet (1977). Because the production, investment, and disinvestment activities are simultaneous, the levels at which the above optimality conditions will be met will be influenced by the investment and disinvestment activities of the firm.

The optimality conditions for the investment/disinvestment decisions of the firm can be summarized as follows:

(4) The firm is optimally organized with respect to investments in durable assets when the net present value of the last unit of durable invested in exceeds its acquisition price while the net present value of the next unit of durable, if it were acquired, would not cover its acquisition cost. The net present value of the durable asset is derived from the services generated in each production period over the economic life of the durable plus the final salvage value.

(5) The firm is optimally organized with respect to disinvestments in currently held durables when the net present value of the last unit of durable disposed of is less than its salvage value while the net present value of the next unit to be disposed of exceeds its salvage value. In meeting these conditions, the firm cannot dispose of more durables than it has in its initial endowment.

These seemingly complicated optimality conditions for investments and disinvestments were necessitated by the assumption that the durables are "lumpy" and by the treatment of time as a discrete rather than a continuous variable; thus, the marginal conditions in the usual sense could not be derived. Each unit of durable should be accompanied by an inventory statement which indicates its physical condition.

The net present value of the durables used in the investment/disinvestment decisions is derived from the net value of the services generated by the durable over its economic life. In calculating the maximum value in use, the firm controls both the amount of services to generate in each time period and the number of periods or the economic life of the durable. The simultaneous determination of the optimal amounts of services to generate from each durable in each period was specified in condition 3.

The optimal number of periods to use a durable asset is when the change in the salvage value from one time period to the next is greater than the value of the services to be generated during that time period. This determines the point in time beyond which it is not advantageous for the firm to use the durable asset. Thus, the maximum value in use for the durable assets was shown to depend on both the optimal amount of services generated in each production period and the optimal number of production periods.

The optimal organization for the durable assets will depend upon the order in which they are altered. The optimal order for altering durable assets was not determined here.

The production, investment, and disinvestment process as developed here permits the treatment of the durable asset replacement decision as a combined investment and disinvestment decision.

Baquet's work builds on and extends work done by Edwards (1958, 1959), Johnson (1960), and Johnson and Quance (1972) in the area of investment and disinvestment with fixed extraction rates, Idachaba (1972) in the area of production with variable extraction rates, Georgescu-Roegen (1971, 1971a) in the area of durable assets in production, and Perrin (1972) in the area of investments in durables with variable lifetimes.

The fixed extraction rate assumption made by Edwards and Johnson has been removed. This permits the production process to be linked with the investment/disinvestment process--something neither Georgescu-Roegen nor Perrin did explicitly.

By specifying the production process in greater detail, it is possible to identify more precisely the manner in which durable assets enter the production process. The conception of the production process as being vertically integrated permitted this development to move beyond Idachaba's work. However, this development considered only the services from the durables to be inputs in the production of the final outputs, whereas Idachaba considered both the stock of the durable and the services from the durable to be inputs in the final production processes.

As a consequence of considering the production process in this manner, it is possible to identify a cost of production and its composition which is usually overlooked--namely, the user cost of generating services from durable assets. This has important implications for firms which practice marginal cost pricing. Previous analyses would indicate a lower marginal cost than an analysis based on this theory.

A further consequence of the vertically integrated production process is in the area of supply response. The analysis here indicates that firms may expand or contract their supply by using their durable assets either more or less intensely rather than by investing or disinvesting in durable assets.

The theory developed by Baquet has important implications for the supply response model presented by Johnson (1960). With fixed extraction rates and divergent acquisition and salvage prices as assumed by Johnson, the firm's supply curve has discontinuities. The theory developed by Baquet would suggest that, when variable extraction rates are considered, the discontinuities would be reduced or eliminated, since firms could respond to changes in output prices by altering the intensity of use of their durable assets rather than by acquiring additional units of the durables or disposing of units of existing durables. The theory presented in Baquet is consistent with Keynes' (1936) suggestion that aggregate output could be varied without a corresponding change in the levels of productive inputs. The theory provides the micro foundation for Keynes' aggregate response.

In much of Georgescu-Roegen's (1971, 1971a) recent work, he has recognized the difference between the durable stock of fund and the flow of services which can be derived from the stock. In this development, the actual process whereby services are extracted from the stock of durable assets is specified. The services are conceived of as being produced from the stock by using non-durable inputs. The modeling of the service generation process in this manner permitted investments and disinvestments in durable assets to be linked with the production activities of the firm. This more detailed specification of the service generation process extends Georgescu-Roegen's earlier writings on the stock/flow conversion problem.

The determination of the optimal investment and disinvestment strategies for the firm relied on comparing an asset's value in use with its acquisition and salvage prices. For the continual replacement problem, the value of using the durable an additional production period should be compared with the annualized average value in use of the replacement durable. The current durable should be replaced when the new durable's annualized average value in use exceeds the value of using the current durable another production period. In computing the asset's value in use, the optimal amount of services to generate in each time period, the optimal maintenance to perform, and, hence, the optimal number of time periods to use the asset are all considered. In developing the criterion for determining the optimal economic life for durable assets, Perrin's (1972) work is extended to consider the "lumpy" durable, discrete-time case.

The criterion is the same as his: Continue to use the durable until the change in its salvage value offsets the change in its value in use. However, the results here appear to be more complex, because neither continuous time nor perfectly divisible durable assets is assumed. The determination of an asset's value in use was specified in more detail than Perrin did in his work. The issue of the optimal order in which to alter durable assets is still unresolved.

Further Research Needs in the Area of Asset Fixity and Supply Response

Further work on the effects of asset fixity on supply response is needed in the following areas:

1. Empirical research is needed which would examine the supply response implications of the theory developed by Baquet. Further research is needed to empirically test the hypothesis that firms can alter output without altering the levels of durable assets.

2. For those firms which practice marginal cost pricing, empirical research is needed to determine if the inclusion of the user cost of durable assets in the firm's decision framework would significantly alter their decisions regarding optimal output levels.

3. Existing models of asset fixity and supply response do not adequately treat the adjustment process over time as firms adjust to changes in their economic environment. Empirical research is needed to explore the consequences through time of following the optimality conditions as developed in Baquet. It is highly probable that this type of analysis will need to be related to the cost of adjustment literature discussed elsewhere in this report.

4. Because the decisions regarding investment and disinvestment in durable assets require information from future time periods, there is an element of uncertainty which exists. Relaxing the assumption of perfect knowledge would be a natural extension of Baquet's work. Including uncertainty would permit a more accurate modeling of the firm's decision environment.

Effects of Uncertainty on Supply Response

The above discussion indicates the need to know various types of information about factors which the managers of electric power generating firms cannot control. This results in a large degree of uncertainty. Relatively little work has been done on the effects of uncertainty on supply responses. In this section, we identify alternative sources of uncertainty.

As indicated in EPRI EA-586-SR, An Overview of the Economic Theory of Uncertainty and Its Implications for Energy Supply, most of the uncertainties facing the energy sector can be characterized in two broad categories: market or endogenous uncertainties, and exogenous uncertainties. Because the remedies for alleviating these two kinds of uncertainty are likely to be different, it is meaningful to distinguish between them.

Exogenous Uncertainty

Exogenous uncertainties exist regardless of the market or legal structure. For the energy sector, the following sources of exogenous uncertainty exist:

1. Geological Uncertainty

   There is uncertainty about the amount and location of deposits of various kinds of natural resources. This uncertainty is likely to lead to slower extraction rates for natural resources. In addition, uncertainty about the cost of extracting natural resources will affect the timing of extraction and, hence, aggregate consumption.

2. Technological Uncertainty

   The development of new techniques is highly uncertain. The timing of development, the costs of development, as well as the cost of operating new techniques are all uncertain. Technological uncertainty has important implications for investment decisions. Should the firm invest in new generating capacity now or should it wait a few years for new technology? This is a very relevant question in the electric power industry.

3. Demand Uncertainty and Variability

Capital-intensive industries such as the energy sector require a long lead time to develop additional capacity. Thus, energy firms must project demand for several years in advance. The general trend in the consumption of energy is not sufficiently stable to provide accurate predictions of the demand for energy.

Endogenous Uncertainty

Endogenous or market uncertainties arise because of actions of other suppliers within the market. The following market uncertainties exist:

1. Investment and Production Decisions of Other Suppliers

The market demand faced by an individual firm is affected by the investment and production decisions made by other firms in the industry. Coordination of investment decisions in the energy sector are of considerable importance because of the long-lived, large capital expenditures involved.

2. Random Pricing Policies

Imperfectly competitive markets can under certain circumstances be shown to be characterized by random prices. Uncertainty associated with the price may affect the attractiveness of different technologies and may introduce, as a consequence, a serious discrepancy between social and private costs.

Other Types of Uncertainty

Uncertainty in other sectors of the economy could have an impact on the prices of inputs and outputs faced by the firm. This price uncertainty is a reflection of exogenous uncertainty in some other sector of the economy. For example, the demand for electricity for heating is affected by the supply of substitutes.

Government-induced uncertainty has a major impact on the energy industry. Changes in governmental regulations, taxes, and the provision of publicly provided goods has a major effect on supply decisions in the energy industry. For example, uncertainty about pollution regulations for particular types of generators may lead to postponement of the installation of any additional capacity until the uncertainty is resolved. This will increase the current price above what it otherwise would have been.

The types of uncertainty identified above imply that the optimal combination of resources for supplying energy is a random variable. The effect that this has on asset fixity and the supply response of firms needs to be investigated more fully.

Models With Risk Components

In this section, we review models that examine how risk affects action choices made by decision makers. We start by defining risk. For our purposes, risk and uncertainty are used synonymously to describe an environment in which any one of several possible outcomes may result from an action choice. Thus, risk requires that decision makers determine their preferred strategy by considering their preferences (utilities) for each possible outcome as well as the likelihood of each outcome occurring as a result of an action choice.

Most would agree that action choices are made in the presence of risk, and economists now are paying more attention to risk in their models. However, decision models with risk components are usually less general than certainty models, which sometimes reduces their usefulness. To illustrate, consider the first-order conditions for determining optimal inputs $x_1, ..., x_n$ to be used in a productive process $f(x_1, ..., x_n)$ for a producer who values possible outcomes of the production process according to utility function U. Moreover, assume he operates in a purely competitive market where output price is p_y and where inputs cost $p_1, ... p_n$. We assume $U' > 0$ and that he chooses inputs so as to maximize the function in (2.2).

$$(2.2) \qquad U[(p_y f(x_1,...,x_n) - \sum_{i=1}^{n} p_i x_i]$$

For the j-th input, the first-order condition requires that:

$$(2.3) \qquad U'(p_y f_j - p_j) = 0$$

or, since U' does not equal zero:

$$(2.4) \qquad p_y f_j = p_j$$

Equation (2.4) is the familiar result that requires the marginal value product of the j-th input to equal its factor cost. Interestingly enough, the result holds regardless of the form of U--all decision makers who prefer more to less will make the same choice of output as well as inputs. Now, however, let price p_y be uncertain; that is, let all possible p_y be described by some probability density function $g(p_y)$, and the general first-order conditions no longer obtain. Assume that the decision maker maximizes (2.2) as before, but now under uncertainty.

$$(2.5) \qquad \text{Max}_{p_y} \int U[p_y f(x_1,...,x_n) - \sum_i p_i x_i] \, g(p_y) dp_y$$

Now, to determine the optimal use of the j-th input, the first-order condition requires:

$$(2.6) \qquad \int_{p_y} U'(p_y f_j - p_j) dp_y = 0$$

Equation (2.6) is, unfortunately, not always easy to evaluate and the function U, which varies between individuals, will alter the optimal choice of inputs as well as outputs. Hence, the generality present in certainty models is usually lost under uncertainty.

The presence of risk not only produces results not easy to generalize but also presents empirical challenges not faced under certainty, such as how to estimate U. Although the theoretical justification and procedures for estimating the function U were established with the development of the expected utility hypothesis, considerable debate has emerged over the reliability of expected utility results. Robison and King (1978) have argued that, in the past, economists attempted to estimate single-valued production functions but found the task difficult because of the large number of factors that couldn't be held constant. Then, in order to extricate themselves from the difficulty of evaluating alternate production strategies when output was stochastic, they assumed that one could estimate single-valued utility functions instead. But, for the same reason that production responses could not be described with a single-valued function, utility is also not likely to be represented by a single-valued function.

The non-uniqueness, as well as the difficulty of actually measuring individuals' utility functions, led economists to use efficiency criteria to help identify preferred action choices. That is, by assuming general characteristics about a class of decision makers or about the action choices they face, it is possible to reduce the set of feasible choices to a smaller "efficient" set that contains the expected utility maximizing choice. More importantly, we can often identify the efficient set based only on characteristics of the probability density functions of outcomes from action choices.

The most well-known efficiency criterion is the Markowitz-Tobin expected value-variances (EV) criterion. This efficiency criterion, often used synonymously with portfolio theory, identifies as the efficient set action choices (portfolios) that minimize variance for given levels of expected wealth or maximize expected wealth for given levels of variance. This criterion has been shown to be efficient when either probability distributions are normal or when investors' utility functions are quadratic. As Robison and Brake point out in their review of portfolio theory, it appears to be most widely used as a financial decision model but has also been used in other disciplines in non-financial settings.

A great deal of debate has focused on how useful EV analysis is when distributions are not normal, with the results indicating, as one might expect, that, as higher moments of probability distributions become less important, EV efficient sets approximate those derived under less stringent assumptions (Tsiang, 1972; Feldstein, 1976; and Borch, 1969).

Of course, other efficiency criteria have been developed: first-degree stochastic dominance (FSD), which assumes positive marginal utility; and second-degree stochastic dominance (SSD), which assumes diminishing marginal utility (Hadar and Russell, 1969). These criteria have been theoretically appealing, yet mostly impractical, because they fail to reduce efficient sets to manageable numbers. In Monte Carlo experiments, where efficiency criteria should be helpful, Anderson (1975) obtained an efficient set of 20 from a Monte Carlo experiment of 50 trials. Unfortunately, most Monte Carlo experiments require a much larger number of trials, say 1,000. Then, if the efficient set increased in the same proportion as it did for Anderson, the number of solutions would still be unmanageable.

Identifying efficient sets rather than single-action choices may be justified for still a different reason. Because of the imprecise measure of an individual's utility function U, the difference between the expected utilities of two probability distributions f and g may not be statistically significant. Thus, while the expected utility of f may be greater than the expected utility of g, ($EU_f > EU_g$), the difference may be so small that one should not reject the hypothesis they are equal ($EU_f = EU_g$). In such cases, the evidence does not permit the analyst to recommend a single-action choice, but rather he should recommend an efficient set of choices.

So, the concept of an efficient set is appealing; but finding methods for identifying reasonably small efficient sets has been somewhat of a problem until the recent work by Meyer (1977). He showed how to restrict efficient sets by setting bounds on risk aversion measures developed earlier by Pratt (1964) and Arrow (1964).

Meyer did not, however, establish empirical procedures for implementing his efficient criterion. Empirical results are currently being tested at Michigan State University with promising preliminary results.

Positivistic Risk Models

Recognizing the importance of risk in the selection of optimal action choices, several economists have attempted to account for risk in positivistic models. Behrman (1968), Just (1974), Traill (1978), Ryan (1977), and Lin (1977) have all introduced risk variables into positivistic models to more accurately measure supply responses under uncertainty.

One recent study (Robison and Carman, 1979) has attempted to provide the theoretical base for positivistic risk models. The study derived demand and supply curves for a market in which a risky asset was traded. It also showed that by using a rather general approximation model, aggregate results could be obtained which did not require the assumption that all decision makers possess the same utility function. Moreover, they also showed how their results could allow for different expectations regarding the variance of the random variable.

Normative Risk Models

The study of normative risk models has presented a tremendous challenge simply because of the large number of possible models to consider. To illustrate, consider the simple production process described earlier, where production was described by the function f determined by inputs $x_1,\ldots,x_n$ at a cost of $p_1,\ldots,p_n$ and output price p_y. Profit π then could be written as:

$$(2.7) \qquad \pi = p_y f(x_1,\ldots,x_n) - \sum_{i=1}^{n} p_i x_i$$

Consider, now, all the possible risk models that could result from such a simple model. These might include uncertain output prices, uncertain response relationships between inputs and outputs, uncertain input prices, or uncertain inputs. In all, 15 different models could be developed even for such a simple production process as described by equation (2.7).

Then, even more variants could be developed by allowing for different firm types, choice variables (inputs, outputs, prices), timing of decisions, types of errors (addition or multiplication), or allowing the firm to purchase information. Consider only one of the variants above and its impact on our economic models. Let Θ be a random variable which enters multiplicatively into the expected utility maximizing problem of equation (2.5):

$$(2.8) \qquad \int_{\Theta} U[p_y f(x_1,\ldots,x_n)\,\Theta - \Sigma p_i x_i]\, d\,\Theta.$$

The first-order condition for the j-th input is:

$$(2.9) \qquad p_y f_j \int_{\Theta} U_j \Theta = p_j \int_{\Theta} U_j$$

If, on the other hand, risk is additive, i.e.,

$$(2.10) \qquad \int_{\Theta} U[p_y f(x_1,\ldots,x_n)] - [p_i x_i + \Theta]\, d\,\Theta,$$

then the first-order condition for the j-th input is:

$$(2.11) \qquad p_y f_j = p_i,$$

or the same results that would be obtained by ignoring uncertainty.

Nevertheless, there are some results that hold in specific models. Sandmo (1971) showed that decision makers operating in a competitive market with diminishing marginal utility and who choose quantity before prices are known will operate at smaller output levels than they would under certainty. He also showed the surprising result that fixed costs do influence optimal output levels under uncertainty. This, of course, conflicts with certainty model results in which fixed costs do not influence output levels in the short run as long as variable costs are covered.

Several studies have examined the impact of risk on production when inputs are lumpy and divisible. Hartman (1976), in a different model, and Turnovsky (1973) examined the optimal conditions for a firm that chooses the durable asset before the output price is known and chooses variable inputs after the output price is known. Unfortunately, Hartman's results produced little that could be generalized, except for the special case of a constant elasticity of substitution (CES) production function. Nevertheless, for the CES function, he showed that the risk neutral firm will increase inputs of the variable factor and output as variability of output price increases, while the optimal size of the lumpy asset will depend on the magnitude of the elasticity of substitution between the variable and lumpy inputs.

Input price risk results have been summarized by Gilbert and Stiglitz (1978). They show that, for the homogeneous risk averse firm and allowing for input price risk with input prices not perfectly correlated, output will be reduced. However, if the firm is risk neutral, the opposite occurs: output is increased. Finally, letting the rental of the durable asset be random reduces its optimal size for risk averse firms but has no effect for risk neutral firms.

Other studies have examined the optimal choice of the durable asset or capacity for a monopolist, who can only adjust the variable input after output price is known. Unfortunately, results are somewhat conflicting depending on the assumptions employed (Rothschild and Stiglitz, 1971; and also Holthausen, 1976).

Summary

The energy industry is highly capital-intensive. Decisions regarding investment and/or disinvestment in capital assets are fraught with uncertainty. These two aspects of the energy sector have important implications for the ability of energy-producing firms to respond to changes in their economic and/or legal environment. In this paper, we have discussed the characteristics of the industry which complicate the investment/disinvestment decision process. The MRRD method for evaluating investment decisions was reviewed. It was pointed out that the MRRD method is not very explicit about some of the calculations of costs which must be determined. A more complete treatment of production and investment developed by Baquet was reviewed. It appears that progress could be made by combining the salient features of the MRRD method with the theory developed by Baquet.

The uncertainty faced by energy suppliers was distinguished as being either endogenous or exogenous. The various sources of uncertainty were identified. The effect of uncertainty on the ability of energy firms to respond to changes is not clear. Several models which are relevant to the energy industry were reviewed. A comprehensive model which addresses all the types of uncertainty faced by energy suppliers has not been developed. More work is needed in this area.

CHAPTER III

THE DEPARTMENT OF ENERGY'S MID-RANGE ENERGY FORECASTING SYSTEM

Glenn L. Johnson

The Department of Energy's (DOE) Mid-Range Energy Forecasting System (MREFS) is more than a model. It is a system of models dealing with the entire energy sector of the United States. It deals with the importation of energy in addition to the discovery and exploitation of energy sources. It has components dealing with coal, petroleum, gas, electric utilities, and synthetic fuels. About the only major source of energy not covered is the photosynthetic generation of energy of carbohydrates in agriculture. MREFS can be interfaced with the DRI macro model, albeit poorly. MREFS also models transportation, refining, electricity generation, and the other processes between the acquisition of basic energy feedstocks and their eventual consumption by final consumers. There is also a major demand component in the MREFS which deals with the demand for energy products in different regions of the U.S. This complex of supply and demand models is integrated with a so-called "integrating model." In this paper, our focus is on the utilities component which handles electricity generation. Our focus, however, is on more than just the supply of electricity, because long-term supply responses obviously depend on what happens to demand. This report attempts to evaluate the adequacy of that part of MREFS which deals with electricity supply, demand, and utilization.

The utilities component of MREFS is an aggregative system. Individual utilities are not modeled; instead, the subsector is modeled. In MREFS, a subsector model, such as the electric utilities subsector, models the building of additional plants of various types in order to supply demand, but the individual utility which builds the plants is not modeled. MREFS simply does not contain such detail for individual utilities. In the model, plants are retired in an exogenous way, but there are no endogenous disinvestments except those associated with conversions and retrofits.

It should be noted that MREFS is large and complex. Its components were put together by a wide variety of people not all of whom are still with DOE. As some parts were not documented by their builders, documentation is not uniform. I have relied on Department of Energy (1977), Stobaugh and Yergin (1979, pp. 254-261) and Vanston and Baughman (1979). I believe these documents tell me enough about MREFS for purposes of the EPRI/MSU project.

MREFS is of methodological interest for the EPRI/MSU project in that it produces supply projections. The EPRI/MSU project is concerned with improving supply projections through better modeling of investment/disinvestment processes and the handling of imperfect knowledge. Thus, this chapter focuses on adequacies and inadequacies in modeling of investment/disinvestment and risk and uncertainty.

Many of the inadequacies of MREFS as an aggregate projection system originate in (1) an oversimplified view of how individual utilities respond to price and institutional and technical change; (2) the effect (through time) of the interactions between supply and demand; and (3) a view of how the demand of individual consumers for electricity responds to price and institutional and technical change which is even more oversimplified than the view of the supply responses of individual utilities, this being important on the supply side because of the interactions between supply and demand in the market place.

The Utilities Component of MREFS

In MREFS, a subsector such as the electric power generating utilities is modeled as if it were under the control of a single maximizing decision body. This is not uncommon in linear programming models, such as the utilities component, which are designed to produce macro rather than micro projections. However, the frequency with which this is done does not make it any more accurate. In the particular case of the electric utilities which are regulated by a wide variety of governmental units and, hence, are by nature locally monopolistic, it is unrealistic to assume that they act as if they were under the control of a single rational decision-making unit.

The linear programming model forecasts the output of three kinds of electricity (base, intermediate, and peak) by regions. Changes in output result, in the model, from changes in environmental regulations and from regulations affecting the use of specific kinds of fuel as well as changes in both input and output prices. The model operates from a base period to a projected point in time.

The linear program for the utilities model is a single period LP which models price responses of the electric utility industry in each of ten subregions of the U.S. The responses to prices are in the form of (1) changed rates at which to operate existing plants, (2) the building of new plants, (3) the retrofitting of existing plants in response to environmental regulations, (4) the conversion of existing plants from one kind of fuel to another, and (5) the retirement of plants.

Other parts of MREFS model the effect of prices and various regulations on the supply of fuel to the electric utilities as they affect exploration and development of petroleum, gas, coal, and other energy resources.

In MREFS, the response of the electric utilities in each of ten energy regions to a set of prices is computed independently in "pre-processor" components of the system.

Once the pre-processor determines the regional quantities of electricity associated with different prices, it uses such price/quantity data to construct regional and industry supply curves for electricity. These supply curves are log linear and are fitted to the output of the linear programs solved in the pre-processor. Initially, these log linear supply curves are reversible, continuous and, of course, of constant elasticity.

There is no corresponding underlying detailed analysis on the demand side; instead, econometric demand functions are estimated from price/quantity time series data available for the U.S. and for the ten energy supply regions used by the Department of Energy in analyzing energy problems. These demand functions are also perfectly reversible and continuous.

Both the supply functions for energy generated in the pre-processors and the demand curves (derived from econometric estimating procedures) are fed into an "integrating component". The integrating component is also a linear programming model which produces a set of equilibrium prices and quantities by successive approximation. It accepts supply and demand functions for the different kinds of energy. It also accepts output from components dealing with refining, electricity generation, and transportation of fuels and different forms of energy among the ten energy-consuming regions of the U.S. The integrating model converts the continuous supply and demand functions into step functions in order to facilitate its linear programming computations--this destroys continuity but not reversibility and leaves overall elasticity basically unchanged.

The integrating LP contains a number of upper and lower bounds on adjustments which can be made from base period conditions as well as a number of accounting or "materials balance" constraints. It also contains a number of "avoid activities" which are used in iterating solutions to attain solutions which "clear the market." If the first set of prices results in more consumption than output, the inequality is avoided with the use of these activities. For the utilities, responses are determined on the basis of marginal costs. For consumers, prices are the average total costs of producing electricity because electricity prices are so regulated. After sufficient iterations are performed, the integrating component reduces the use of "avoids" to acceptable limits and the program is regarded as balancing production with consumption at a set of equilibrium prices. Thus, the output is a set of equilibrium prices for different kinds of energy associated with equilibrium quantities produced and consumed, by regions and for the U.S.

It should be stressed that the solution of the integrating component is a single-period solution. Solutions are computed for any year in which the Department of Energy is interested. The program goes from its initial point in time to an equilibrium position at a point in time such as 1980 or 1985, the point in time being defined by assigning values to exogenous variables expected or assumed to prevail then.

It would be possible to obtain projections from MREFS, by years, from any base point in time as far into the future as one dares project MREFS' exogenous variables. This could be done in two ways: (1) from the base period to each year, or (2) from the base year (t) to (t+1), then from (t+1) to (t+2) up to (t+n), the last year of interest. Overall, MREFS is such a large system involving inversion of so many large LP matrices that solving it year by year would be such an expensive, time-consuming process that it is not done often, if at all. Documentation pertaining to MREFS indicates that it has been used mainly to make projections from a base to a single future year. Overall descriptions of MREFS do not refer to auxiliary equations for moving automatically from (t) to (t+1). MREFS does not seem to be a very useful model for tracing out the "time trajectories" of important energy variables because of the complicated matrices which are expensive to invert in on-going solutions. For the same reason, neither is it well adapted to (1) Monte-Carlo analysis of the consequences of incorporating probability distributions, or (2) sensitivity analyses.

Investments, Disinvestments, and Imperfect Knowledge on the Supply Side of the Electric Utilities Model of MREFS

Output responses are forecast under the implied assumption that there is perfect knowledge. To repeat an earlier observation, the forecasts are generally for a point (year) in time and represent change from a base period.

Utility Investments

In the electric utilities model, supply responses originate from changes in the rates at which existing plants operate, from investments in new plants and equipment, and from disinvesting in existing plants. This section concentrates on the modeling of investment responses in MREFS.

The linear programs model investment responses mainly from a base period described in terms of existing plants to some future projected point in time defined in terms of the projected values of exogenous variables at that point. Investments considered include the building of new plants, the retrofitting of existing plants, and the conversion of existing plants from one kind of fuel to another. When the price of electricity is high enough, the linear program builds new plants. Retrofitting investments, on the other hand, are likely to result mainly from scenarios which impose regulations requiring the retrofitting of plants to include additional pollution control devices, safety measures, etc. The conversion of existing plants from the use of one kind of fuel to another may occur as the result of either (1) scenarios which require that such changes take place, or (2) changes in the relative prices of fuel which make it advantageous to convert the plants.

The documentation examined does not indicate that the activities involving additional investments have been modeled so as to recognize the interdependence of activities resulting from the conversion of stocks into flows and the associated costs and prices. Recognition of such interdependence among the activities of an LP would probably require moving to another technique and a more general methodology. The EPRI/MSU project is considering such techniques and methodologies.

Lack of Attention to Time and Imperfect Knowledge

The energy industry operates in a changing environment poorly understood and known by the public and private decision makers who control supply and demand responses. Uncertainty has increased greatly in recent years as a result of resource exhaustion, OPEC actions, changing environmental health and energy regulations, price controls, political instability, and threatened or actual military operations. The electric utilities subsector is affected by these uncertainties as much as any other subsector of the energy sector.

Such uncertainty causes mistakes in the investment decisions of the electric utilities. These mistakes are made year by year, time period by time period. Once an investment mistake is made in the electric utilities industry, it influences future investments. Complete disinvestment in a plant is seldom feasible--instead, a utility has to endure its consequences for a long time. These consequences influence subsequent investment decisions. This is true even if demand is expanding so rapidly that overinvestment mistakes do not have to be liquidated.

The above line of reasoning raises questions about a MREFS projection which goes directly from, say, 1974 to 1985 without attention to the sequences of decisions and mistakes made year by year in the intervening 10 years. Now that we have experienced half of the intervening time period, the dangers of solving a single-period LP in 1974 for 1985 are all too obvious. The prospects for a single LP solution in 1980 for 1990 now look even dimmer.

Utility Disinvestments

The MREFS system devotes very little attention to the modeling of disinvestment by utilities. The supply functions generated in the pre-processor are reversible but are not asked to model reversibilities.

Three kinds of disinvestments are considered. There is an exogenously determined rate at which existing plants are retired due to obsolescence and wear. This rate does not respond to either product or input price. Secondly, each conversion of an existing plant from one kind of fuel to another implies a disinvestment in the equipment and capital specialized in utilizing the kind of fuel abandoned. To the extent that the model forecasts plant conversions due to changes in the relative prices of fuel, these investments and disinvestments are reversible, though the model is not used to consider such cases. Conversions due to changes in regulations in scenarios are not reversible. Even when scenarios are changed, disinvestments are not considered between scenarios; instead, the model is run twice, with the runs for both scenarios originating at a common point in time. Thirdly, in connection with some retrofits, there may be disinvestments in equipment and capital, but, inasmuch as retrofits result largely from variations in regulations from scenario to scenario, these disinvestments are exogenously determined.

It should be stressed again that the linear programs forecast changes from a base period to a selected point in time. No attention is given in the model to the dynamics and sequences of events from that base period to the final point. The transition from a base period to a projected period is modeled under the implicit assumption of perfect knowledge.

Investments, Disinvestments, and Imperfect Knowledge on the Demand Side of the Electric Utilities Model of MREFS

Though the EPRI/MSU interest is mainly in supply projections, interactions between supply and demand make it necessary to consider demand. The demand components of MREFS are traditional. Each demand function, as initially computed econometrically, is continuous, log linear, and has constant elasticity. The demand function for each energy product contains cross elasticities of that energy product with each other energy product considered in the model. The continuous log linear demand functions are converted to step functions for use in the integrating model along with the step supply functions. The integrating model is a linear programming model. Sets of equilibrium prices and quantities are arrived at by successive iterations.

Consumer Investments

No provision is made for capital investments on the part of consumers to shift demand curves for electricity or for any of the other energy products considered in MREFS. For example, if electric cars were to be substituted for gas-driven vehicles, there would be no endogeneous way of indicating when such investments would occur.

Clearly, the utilities component of MREFS would not be able to model the demand consequences of such shifts. Similarly, current disinvestments of consumers in large cars are not reflected in the cross elasticities of demand between gasoline and electricity, nor would the substantial current investment of householders in natural gas conversion burners and new furnaces be reflected in the cross elasticities of demand for fuel oil and natural gas. The absence of components which model consumer investments and disinvestments in equipment using different forms of energy means that the influence of these investments on energy feedstocks for the electric utility industry is not taken into account in MREFS, which, in turn, raises questions about the reliability of supply projections for electricity.

Consumer Disinvestments

Similarly, there is no attention in MREFS to endogenous determination of when consumer disinvestments are likely to occur. Disinvestments and early retirement of inefficient air conditioners, electric home heating equipment, and other electricity-using appliances are not modeled. The exception to this statement would be in the extent to which the traditional log linear demand functions for electricity reflect disinvestments with price increases and investments with price decreases. It is not thought that this exception is important enough to offset the deficiencies of the MREFS demand model for projecting the consequences of major changes in electricity prices such as have been occurring and are likely to occur before 1985 and 1995. There will be major disinvestments in consumer goods and in the rates at which consumers extract services from their durables which will affect both own and cross elasticities of demand for electricity in ways not reflected in the data used to generate the elasticity estimates.

Lack of Attention to Time and Imperfect Knowledge

The demand side of MREFS, like the supply side, implicitly assumes imperfect knowledge and does not take risk and uncertainty into account. In view of the increased uncertainty associated with foreign supplies of energy and the activities of regulatory agencies, this seems to be an important shortcoming of the demand side of MREFS and, more specifically, of the utility component of MREFS. When knowledge is imperfect, substantial investment and disinvestment mistakes are made by energy consumers. These mistakes occur at different points in time. The fact that the MREFS model solves for changes occurring between points in time (without attention to intervening sequences of mistakes and consequent attempts of consumers to minimize their losses on such mistakes) raises grave questions as to the reliability of MREFS demand components. Even if MREFS, as now formulated, were operated from time period to time period (sequentially, by years), the time trajectories would not be expected to be accurate in this respect.

Other Evaluations of the Electric Utilities Model of MREFS

The other evaluations to be considered are those by Vanston and Baughman (1979) and by Stobaugh and Yergin (1979). While done with somewhat different criteria in mind, these two evaluations are basically consistent with the above.

The Vanston/Baughman (VB) Evaluation

VB make the following three points about the electric utilities component of MREFS.

1. ". . . the model will forecast behavior very poorly if unconstrained. . . Future expansions and plant utilization possibilities are entered in a tightly constrained set of data."

2. "The model is used simply to translate one set of prices (fuel and capital) into another (electricity)."

3. "The electric utilities representation in the LP is the simplest possible."

VB's first point is consistent with the stress in this chapter on the oversimplification of the model with respect to investment and disinvestment by utilities and consumers. With respect to the second point, it is feared that the translation of prices may be of questionable accuracy not so much because of the oversimplifications in the electric utilities component as in other parts of MREFS; the sequences of consumer and producer investments and disinvestments under uncertainty which affect own and cross elasticities are not modeled in a way which should be expected to feed the utilities component the proper fuel and capital prices to be translated into electricity prices. It should be noted that, while the models are simple conceptually and because they are simple abstractions of reality, they are computationally complex, i.e., they require complex, time-consuming matrix inversions and iterations which increase costs and constrain use of the model.

The Stobaugh/Yergin (SY) Evaluation

This evaluation of the entire PIES (later MREFS) contains little specific to the electric utilities component. However, the overview provided has value for MSU/EPRI purposes as indicated in the following.

> PIE-74, later MREFS, as did all other energy models formulated at that time, contained the problems associated with . . . the assumptions that range of prior experience is relevant and that the processes are reversible. These problems . . . led to the determination of several elasticities that were contrary to what theory would presume. For example, the equations which were derived by fitting curves to historical data, showed that in the household-commercial sector the demand for natural gas would fall as the price of oil increased. According to theory, however, the opposite should be true--that is, the demand for natural gas should rise because higher oil prices should encourage consumers to switch to natural gas. Many other such unexpected elasticities were also obtained from the model. Also, some elasticities declined with time, whereas one would expect responsiveness (and hence, elasticity) to increase over time. Furthermore, some elasticities were considered too high. Consequently, many judgmental modifications and adjustments were made. Such adjustments, observed the General Accounting Office, 'raise questions regarding the accuracy of all elasticities developed by the system.'

This, of course, is similar to the concerns expressed above in connection with the VB evaluation. While the electric utilities model may be a good translator of fuel and capital costs into electricity prices, good translation does not offset the effects of inputting incorrect fuel and capital costs.

Some Conclusions

It is concluded that the electric utilities component of MREFS suffers from inadequate:

1. conceptualization of investments and disinvestments;
2. attention to changes in the rate at which services are extracted from fixed investments in response to changes in prices associated with technical (including resource availability) and institutional and human change; and
3. attention to sequences of events between base periods and projection points, particularly those resulting from investment and disinvestment mistakes made by both the electric utilities and their consumers.

It appears, therefore, that the objectives of the MSU/EPRI project are important in improving our ability to forecast electricity supplies. In this connection, it is gratifying to note that the Office of Applied Analysis, Energy Information Administration, of the DOE, announced on December 5, 1979, that they anticipate granting contracts on:

1. Refinery Investment Analysis

The purpose of this procurement is to provide a dynamic methodology which will look at the optimum time path of investment given assumed product price levels that the refiner expects to see over time and will permit the analysis of economic, technological, environmental, or institutional factors which may affect the industry's ability or willingness to invest in new or modified capacity.

2. Investment Decisions in the Face of Uncertainty

The purpose of this procurement is to investigate the investment decisions of firms made in the face of uncertainty. This would provide a more realistic assessment of probable energy capital investments by firms, the integration of these effects into the various financial models, and an assessment of the possible impact of various types of federal subsidies on the investment decisions of firms engaged in new technologies. Such a study would look at the theoretical impacts of uncertainty on firm behavior, and at how effective various types of federal subsidies have been in inducing investments in areas with high real or perceived risks. The contract would evaluate how different sized firms deal with uncertainty and the possible methods or approaches which could be used to overcome this uncertainty.

CHAPTER IV

INVESTMENT/DISINVESTMENT AND USE OF DURABLES: AN ANALYTIC FRAMEWORK

Lindon J. Robison

Introduction

Electric utilities have the responsibility of meeting a time-varying demand for their product, a demand that often depends on factors outside their control, such as weather. Their ability to meet this time-varying demand is determined in advance by the cumulative results of past investment/disinvestment, use, and maintenance decisions. These decisions, in effect, determine the capital stock of durables from which services to meet demand can be extracted. The level of capital stock of durables maintained by the utilities to meet demand depends on the probability with which they desire to meet customer demands. A higher probability of meeting peak demand requires a larger capital stock; a lower probability of meeting peak demand requires a smaller capital stock.

In all of these decisions, utilities face the requirements of meeting operating costs and earning an acceptable return for investors. The returns actually earned depend on at least two things: the quantity of their product sold and the rate at which customers are charged. The quantity demanded, of course, depends on the need for their product. The rate depends, at least in part, on regulatory commissions.

In determining an appropriate rate for utilities to charge their customers, the regulatory commissions establish a rate base. This rate base is the utilities' capital stock on which they are allowed to earn a return. The rate actually charged customers by utilities is set in such a way as to allow them to earn an acceptable return on their rate base and pay operating costs.

In determining which capital items are to be included in the rate base, the commission applies a least-cost-to-the-customer criterion--that is, the utilities should acquire capital items in such a manner that rates charged customers are minimized. Capital acquisitions that, in the view of the commission, depart from this criterion have a higher probability of not being allowed in the utilities' rate base. As a result, utilities have an incentive to produce in a least-cost manner.

So how do utilities organize production to meet customer demands for an energy product in a least-cost manner? Obviously, capital investment/disinvestment decisions have an important impact on costs, but capital investment/disinvestment decisions cannot be made independently of decisions regarding the use of capital items and decisions regarding the purchase and use of non-durable inputs (inputs used up in a single period).

Both the availability and costs of non-durable inputs and the amount of services extracted from the capital stock affect both the optimal size and the optimal time to disinvest/invest in capital items. Building a conceptual framework that includes in it these interdependent relationships is the goal of this chapter.

In what follows, an analytic framework is developed for solving capital investment/disinvestment and use problems. In the process, this chapter identifies costs associated with capital or durable ownership and use; specifies the optimal conditions for replacement; deduces a new theorem which permits the valuation of durables' services; and, finally, solves a simple example to illustrate the analytic model. Before we proceed further, however, we define two classes of assets: durable assets and non-durable assets. Because several new concepts and definitions will be introduced, a glossary is included at the end of this chapter to aid the reader.

Durables versus Non-durables

For an arbitrarily defined period, non-durable assets are used up, i.e., do not exist in the same form after a single period. Durable assets are not used up; they exist in nearly the same form for more than one time period. Durables may be either divisible or lumpy; or, they may be reversible (positive salvage value) or irreversible (zero or negative salvage value) investments. Their distinguishing characteristic is not divisibility or reversibility; their distinguishing feature is their existence beyond an arbitrarily defined time period.

If non-durable assets do not have a life beyond a single time period, then their costs are the costs associated with their acquisition and use. (It might also be added that if the non-durable is used up in a single period--one need not ask when to disinvest.) If durable assets, on the other hand, have a life beyond a single period, then there are costs associated with their acquisition and use plus those costs associated with holding an asset inventory over time. A third cost category results from the impacts of use on future inventory costs.

Accounting for these three categories of costs is an important part of the construction of an analytic model that prescribes investment/disinvestment and use of durable assets. So we begin on the cost side of the model and examine in some detail each of the three major cost categories: (1) those current period costs incurred because of changes in the capacity of the asset to deliver services, either as a result of use or the passage of time--capacity costs; (2) those costs that occur as a result of holding an inventory of extractable services over time--inventory costs; and (3) those future-period costs (benefits) resulting from current-period use decisions--indirect capacity costs.

Clearly, the distinction between durables and non-durables depends solely on whether or not the asset exists over an arbitrarily defined time period. The shorter the time period, the larger the class of durables relative to non-durables. If the period is a month, gasoline purchased for a passenger car on the first day of the month is a non-durable asset. If, however, the period is a day, the gasoline is a (divisible-reversible) durable. Finally, by making the arbitrarily defined time period shorter and shorter, all assets become durable. The benefit of our distinction between durables and non-durables is to allow the decision maker himself to determine which assets are durable based on his relevant planning period.

Having distinguished between durables and non-durables, and between capacity, inventory, and indirect capacity costs, we are now prepared to launch an investigation into each cost category associated with investing/disinvesting and using durables. We begin by identifying capacity costs associated with durables.

Capacity Costs Associated With Durable Assets

To analyze systematically the capacity costs associated with durable assets, we divide capacity costs into three categories: (1) those that occur as a result of use, called direct user costs; (2) those that occur as a result of time, called capacity time costs; and (3) those that occur as a result of maintenance, called maintenance costs.

Direct User Cost and Capacity Time Cost

Direct user cost is the replacement cost of an asset used up. In the case of non-durable assets used up in a single time period, the direct user cost equals its acquisition price. This price, a cost to the firm, is a use charge for converting the asset from an input to an output through a production process. Since its form is changed and is not reversible, its salvage price is zero. There is a similar cost associated with using a durable asset in a production process. This cost, like the acquisition price of non-durable assets used up, reflects the replacement cost of that portion of the durable's capacity "used up" in the current period. In some cases, there is also a loss in the capacity of a durable to delivery services as a result of the passage of time alone; that is, time alone determines the rate at which the durable's capacity is used up. Examples of such time-related depreciations are: roofs or painted exteriors on buildings; miles that a tire can be driven; the finish on a car; etc.

Measuring the value of the durable's capacity used up is more complicated than measuring the value of non-durable assets used up because prices change over time. In addition, the quality of the durable may be altered as a result of time, maintenance, and use. How to measure and value these changes in the durable's capacity is the problem now faced--a problem further complicated by the fact that physical appearances may not always depict changes that have occurred in the durable's capacity.

Consider some examples of capacity changes. The change in the number of gallons of gasoline in a gas tank is an accurate measure of the amount of that durable used up. Also, the miles of service extracted from a car may be a good indication of the service it has provided--but not always. A car driven at 50 miles per hour for 100 miles has not had its capacity to deliver transportation services affected in the same way as a car driven at 100 miles per hour for one hour. Complications like these, as well as the effect of time on a durable's capacity to deliver services require careful attention to the measurement problem.

In measuring costs associated with using up a durable, or the cost of extracting services from a durable, the most relevant question becomes: what is used up? A related question is: what is purchased that can be used up? Consider, for example, the purchase of a consumer durable, a passenger car. The car is purchased primarily because it provides transportation services--it moves people and things from one place to another. Its ability to deliver these services can be measured in at least two ways. The first may be the total miles of transportation services expected to be delivered. This capacity may be measured by the car's expected odometer reading when parked in its final resting place in an auto salvage yard. The second capacity measure may be the car's ability to deliver transportation services during a fixed and shorter time interval, say one hour. The miles of transportation services a car delivers in an hour may vary between 0 and 90 miles. The second capacity measure, then, is not unique but may be a range of possible service extraction rates.

We refer to the first capacity measure as the durable's lifetime capacity. The second measure or range, we refer to as the durable's operating capacities. These two measures, lifetime and operating capacities, are, of course, related. The lifetime capacity of a durable may depend on the operating capacities used to extract durable services. A car driven at an operating capacity of 55 miles per hour may have a lifetime capacity of 100,000 miles of service before parked in an auto salvage yard. Meanwhile, the same car driven at an operating capacity of 90 miles per hour may produce only 50,000 miles of service, not to mention the reduced lifetime capacity of the car's operator. The operating capacity which maximizes the lifetime capacity of the durable is defined as the durable's rated capacity.

Having stated how the operating capacity used to extract services affects lifetime capacity, it should also be pointed out that the operating capacity used may also influence operating capacities available in the future. A car driven 90 miles per hour for 50,000 miles does not have the same operating capacities to deliver future services as a similar car driven at 50 miles per hour for 50,000 miles. It may require more frequent trips to the garage or, even worse, its services may be interrupted unexpectedly. As a result, most used car buyers have some concern for both the units of service extracted (as measured by the odometer reading) as well as the operating capacities at which these services were extracted.

More factors than the operating capacities used to extract services affect a durable's remaining lifetime capacity. External factors may include: (1) conditions under which services are extracted, e.g., Arizona versus Michigan weather; (2) maintenance, both scheduled and unscheduled; (3) quality of inputs used in combination with the durable; and (4) time interval over which the services are extracted, to name a few. For the moment, we focus on the relationship between the operating capacity used to extract services and time in determining the durable's lifetime capacity.

We have illustrated how operating capacities or use influence lifetime capacities. Consider an illustration of how time may also influence lifetime capacity. A set of tires is sold with a guaranteed lifetime capacity of 40,000 miles or 5 years, whichever comes first. Suppose these tires are placed on a car in which services extracted per year are minimal. As a result, long before 40,000 miles of tire services have been extracted, chemical reactions between the tire and the atmosphere reduce the tire's lifetime capacity to something less than 40,000 miles. Recognizing this fact, tire manufacturers limit the length of time their lifetime capacity guarantee applies.

Since time is independent while the operating capacity used is under the control of the decision maker, we could express lifetime capacity as a function of the operating capacity used and time. This relationship will provide us the measure of what is used up when services are extracted from a durable. Determining what is used up is what we require to eventually obtain our "capacity cost" measures. Having determined such a relationship, we might also ask: what choice of operating capacity would maximize the lifetime capacity of the durable? Having determined such an operating capacity, we define it as the durable's rated capacity to distinguish it from other operating capacities.

The relationship between operating capacity, lifetime capacity, and rated capacity can be illustrated graphically. Let the vertical axis in Figure 4.1 represent losses in lifetime capacity and let the horizontal axis be the operating capacities that can be used to extract services. If there is no loss in lifetime capacity as a result of time, and if the marginal loss in lifetime capacity associated with use is constant, then services extracted equal lifetime capacity lost regardless of the choice of operating capacity, a relationship

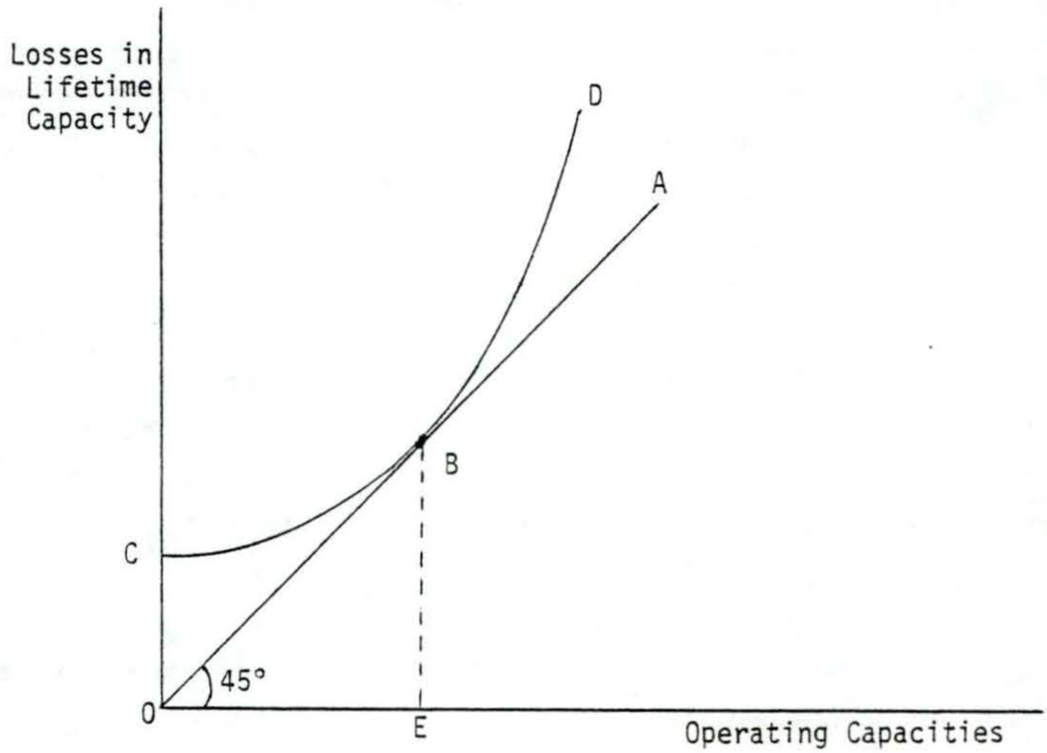


Figure 4.1

Relationship Between the Rate of Services Extracted
From a Durable and Losses in Lifetime Capacity

depicted by the 45-degree line OA in Figure 4.1. Because services extracted equal losses in lifetime capacity of services at the rated capacity, the slope of any functional relationship between the two capacity measures cannot be less than one.

To suggest that losses in lifetime capacity equal services extracted at more than one operating capacity implies there is no unique operating capacity that maximizes lifetime capacity; that is, services may be extracted from the durable at different operating rates and still obtain the same lifetime capacity. We define durables with various rated capacities as "flexible," while those with a unique rated capacity are defined as "inflexible" durables.

To exemplify inflexible durables, consider one with increasing marginal losses in lifetime capacity, represented by curve CBD in Figure 4.1 and loss levels resulting strictly from time of OC. Whenever marginal losses in lifetime capacity associated with use are increasing and there are positive time costs, then there will always be some unique operating capacity, a rated capacity that maximizes lifetime capacity--and the durable will be inflexible.

The rated capacity can be found graphically for the inflexible durable CBD in Figure 4.1 at its point of tangency with the 45-degree line emanating from the origin. For the inflexible durable described in Figure 4.1, this tangency occurs at operating capacity OE. The slope of this linear tangent line is the minimum average loss in lifetime capacity, equal to total losses in lifetime capacity divided by the operating capacity. Obviously, the slope of curve CBD at its rated capacity must equal one.

The rated capacity, OE, represents, then, the operating capacity that minimizes average losses in lifetime capacity. To demonstrate this result, we graph average losses in lifetime capacity at various operating capacities. Consider first the flexible durable. For the flexible durable, there is no one service extraction rate that minimizes average losses of lifetime capacity. As a result, average losses are the same, regardless of the operating capacity used to extract services. This result is obtained graphically by dividing total losses (line OA in Figure 4.1) by services extracted or the operating capacity used. The result of this division is the constant average loss line AA' in Figure 4.2.

For the inflexible durable (with total losses described by curve CBD in Figure 4.1), the average loss curve decreases as operating capacity used to extract services increases at least up to the rated capacity OE; then the average loss curve increases. At first, the average loss curve decreases with increases in operating capacity used to extract services because the average loss in capacity due to time decreases as time costs are averaged over more and more units of service. However, as operating capacities used to extract services increase beyond the rated capacity OE, increasing marginal losses associated

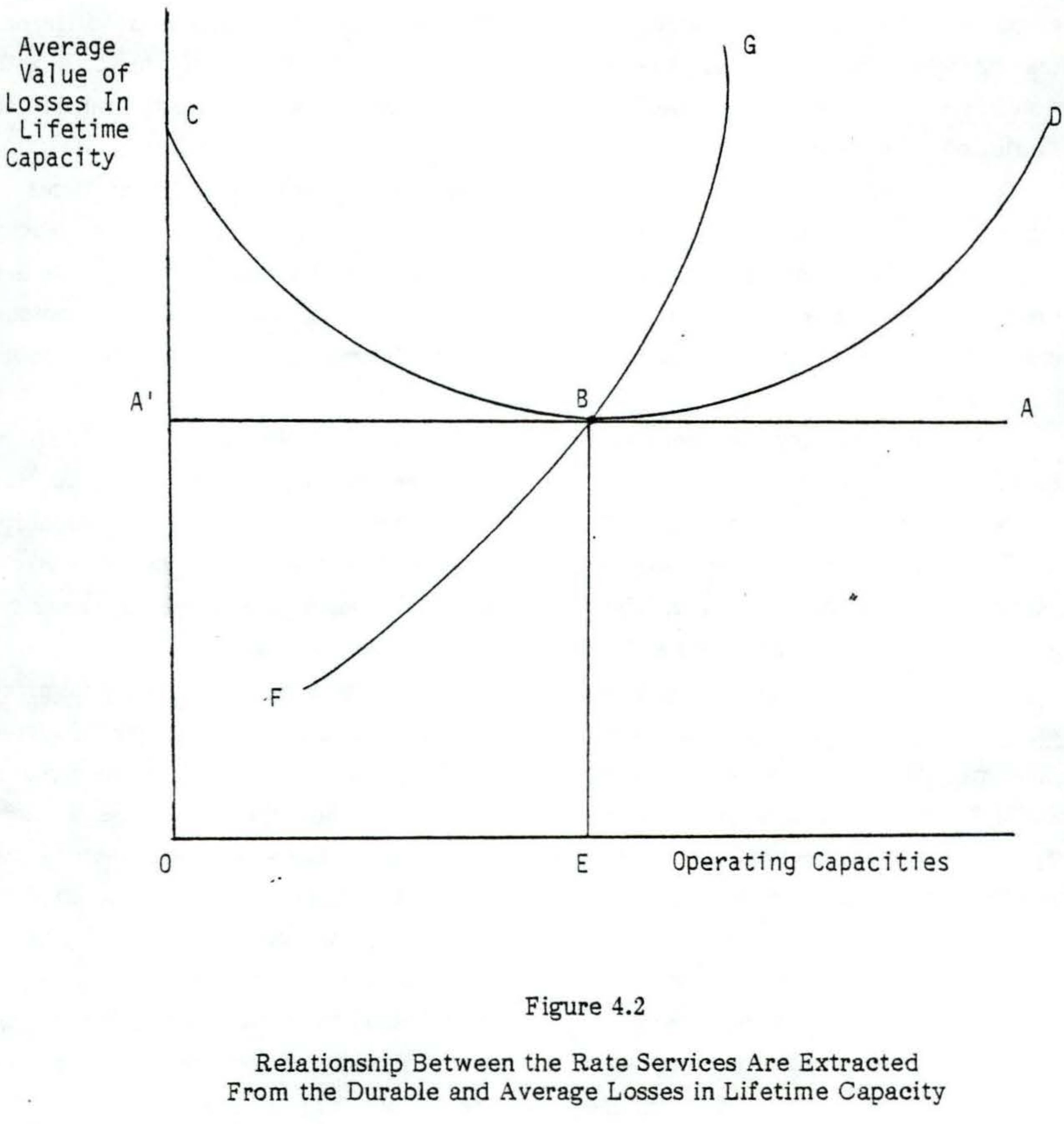


Figure 4.2

Relationship Between the Rate Services Are Extracted
From the Durable and Average Losses in Lifetime Capacity

with use more than offset the decrease in average losses due to time. The graphical result is a U-shaped average loss curve CBD in Figure 4.2, which corresponds to the total loss curve CBD in Figure 4.1.

The relationship between average losses in lifetime capacity and services extracted is an important matter for utilities. A flat average loss curve indicates that no particular cost advantage is associated with a specific output. Thus, a flexible durable is more desirable if output is variable, as it is for utilities trying to meet time varying demands. Inflexible durables, on the other hand, do incur greater costs by producing at rates other than their rated capacity. On the other hand, inflexible durables may have a smaller average loss at their rated capacity than attainable from flexible durables. (See also Gilbert, Newberry, and Stiglitz, 1978.)

Now, having identified the durable's rated capacity, we are in a position to identify lifetime capacity of the durable. Assume the durable described by curve CBD can be operated for t* years at service extraction rate OE and that time capacity losses per year equal OC. Then the durable's lifetime capacity becomes the sum of the losses in capacity due to time (t* X OC) plus losses due to use (t* X OE). The lifetime capacity F* of the durable in Figure 4.1 could then be represented as:

(4.1) $F^* = t^* (OE + OC)$

As pointed out already, inflexible durables have only one rated capacity. To be completely flexible, a durable must have no loss in lifetime capacity due to time and have a constant marginal loss in lifetime capacity associated with use. Then, extracting services at higher operating rates does not result in a greater average loss in lifetime capacity. If the durable has either positive losses in lifetime capacity due to time, or if the marginal loss in lifetime capacity varies with operating capacities, then the durable has a unique rated capacity. Of course, it is possible to conceive of a pathological case where the marginal losses of lifetime capacity increase or decrease according to operating capacities, in which case more than one rated capacity exists. This would be an unusual case, however, and one we will not consider.

From equation (4.1), assuming we know service extraction rates and lifetime capacity, it becomes a simple matter to solve for the durable's calendar life. The calendar life is the ratio of lifetime capacity to the durable's rated capacity plus average time loss per period. For the flexible durable with lifetime capacity F and operating capacity OE, it is:

(4.2a) $t = F/OE$

For the inflexible durable, also with lifetime capacity F*, it is:

(4.2b) $t^* = F^*/(OE + OC)$

The relationship between lifetime capacity, operating capacities, rated capacities and lifetime capacity losses due to time are often predetermined engineering results. As such, this leaves little for the economist to prescribe, except perhaps as to the optimal time to disinvest. On the other hand, if consulted in advance, the economist could provide useful suggestions as to the optimal relationship between services extracted and losses in lifetime capacity given engineering design constraints. Of course, good engineering results and good economics would result in equivalent design.

Converting Physical Loss Measures to Money Losses

The loss measures described thus far have been physical, dependent upon the durable and the type of service it can deliver. The next step is to value in dollars the cost of using up the durable or altering its service capacity through time.

The acquisition price, if the durable is being purchased, or the salvage price, if the durable is already owned by the firm (hereafter referred to as the durable's price), reflects the present value of services expected from the durable. Since many units of potential services are acquired, and each unit of service may not be valued equally, the price reflects a composite price for all services. It may also reflect the desire to acquire a certain operating capacity, as well as a lifetime capacity measure. The durable's price can also be viewed as an average price per unit of lifetime service, $\overline{P}$, multiplied by the units of lifetime services available, F^*. If the relationship between the F^{th} unit of service extracted and the acquisiton price of that unit of service is given by $P(F)$, then the average price for all units of service is given by:

$$(4.3) \qquad \overline{P} = (1/F^*) \int_0^{F^*} P(F)dF$$

while the acquisition price of the durable is $\overline{P}F^*$. Equation (4.3) says that the average price $\overline{P}$ of all units of service is the sum of what is paid for each unit dF divided by the total number of units F^*.

Equation (4.3) represents an equilibrium price. It is a price that suppliers of the durable expect will at least cover costs of production and that purchasers of the durable expect will be equal to or exceeded by the value of services extracted from the durable. Consider how a purchaser of the durable might decide on whether the value of the durable exceeds its acquisition price.

To find the maximum price a buyer would offer for a durable, we adopt some simplifying assumptions. For example, assume that each unit of service extracted from the durable is valued equally at rate R. In addition, assume that the durable is the inflexible one depicted by curve CBD in Figure 4.2; moreover, assume that in each period services are extracted at the durable's rated capacity OE. Extracting services at this rate implies a calendar life of t*. Finally, because of opportunity losses due to the passage of time, let future benefits be discounted by rate r. Then, summing the discounted present value of returns from the durable at time period t_o, we obtain equation (4.4), the maximum price P(t) a firm would be willing to pay to acquire the durable:

$$(4.4) \qquad P(t_o) = \int_o^{t^*-t_o} R(OE)e^{-rt}dt$$

As services from the durable are used up and the lifetime capacity is reduced, the value of the durable is similarly reduced. We graph the maximum bid price, P(t), against time as the durable's lifetime capacity is used up. This relationship is given in Figure 4.3. The rate of change in the durable's price over time is found by taking the derivative of $P(t_o)$ as the remaining calendar life (t^*-t_o) is reduced.

$$(4.5) \qquad d\,P(t_o)/dt_o = -R(OE)e^{-r(t^*-t_o)} < 0$$

That is, the durable's price is reduced in each period by the returns in the last period discounted to the present. Further, as Figure 4.3 illustrates, the price relationship is convex. The convex relationship holds because, as the calendar life is shortened, the last period of services available is closer to current time t_o.

Having determined $P(t_o)$, a maximum bid price, the firm compares it to the acquisition price $\bar{P}F^*$ and acquires it if $P(t_o)$ is greater. Then, as services are extracted from the durable, the value in use price $P(t_o)$ is continually compared to the market price, and the durable is retained by the firm as long as value in use exceeds the market or salvage price. Meanwhile, the change in the durable's salvage price associated with using up the durable reflects the cost, a direct user cost and time capacity cost, incurred by the firm in order to extract services from the durable.

Again, if the market valued each potential unit of service of the durable equally, the average dollar cost of using services from the durable could be obtained by multiplying the average loss in capacity curves in Figure 4.2 by $\bar{P}$.

The cost of using the durable, then, is the units of capacity used up, valued by the market or salvage price. Pricing the durable's capacity used up allows us now to relate operating and lifetime capacity in terms of costs rather than physical units of capacity lost because of time and use. Formulating this relationship is the precursor to developing our economic model.

Maintenance Costs

Our third capacity cost is maintenance cost. It represents still another way the capacity of the durable may be altered, as a result of either planned or unplanned

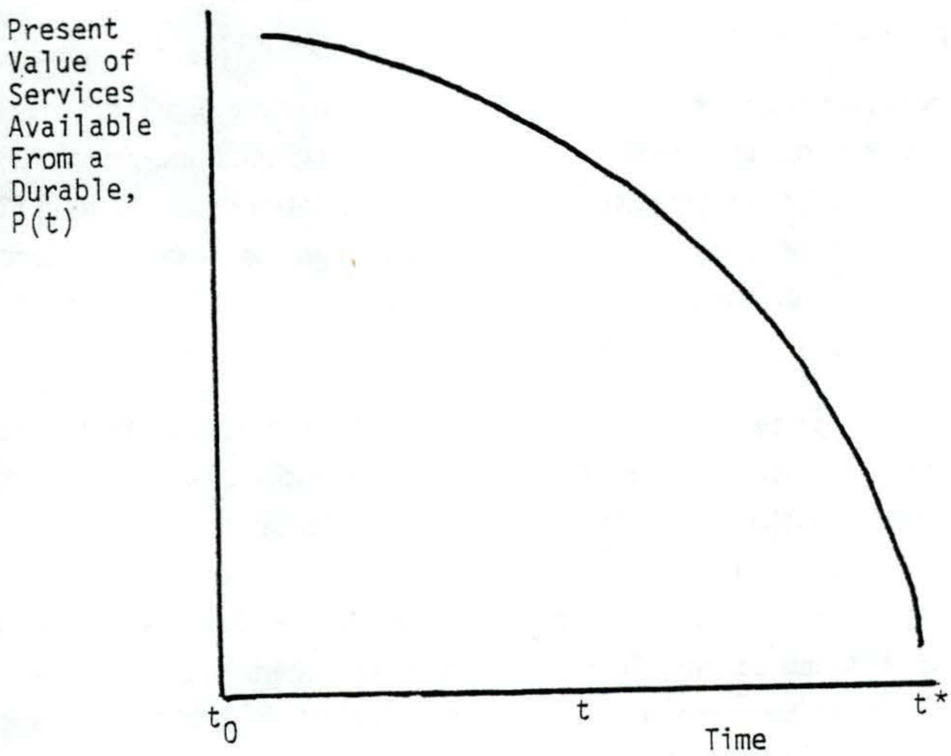


Figure 4.3

The Relationship Between Time and the Present Value of Services Available From a Durable Assuming Services Are Extracted at a Durable's Constant Rated Capacity maintenance. Scheduled maintenance is a cost that is designed to alter the losses in lifetime capacity associated with time and use. Also, because the services derived from maintenance may extend beyond a single time period, maintenance itself may be considered a durable investment. Therefore, the decision to maintain a durable also becomes an application of investment/disinvestment theory.

Again, returning to our car example, oil changes and lubrications are maintenance expenditures designed to reduce the losses in lifetime capacity of the car as services, in the form of passenger miles driven, are extracted. Obviously, maintenance will be performed as long as the value of the lifetime capacity savings exceeds the cost. In fact, the decision to park the auto in the salvage yard is most often the result of a conscious decision that the cost of maintenance exceeds the value of capacity changes resulting from the maintenance. Alternatively, the returns expected from maintenance may be less than the cost because of the shortened remaining lifetime capacity of the durable.

One might also view the value of maintenance investments to be inversely related to lifetime capacity. To perform regular maintenance on a new car with 100,000 miles of expected lifetime capacity is usually considered a good investment because of the large potential loss from failure to maintain it. On the other hand, there is only a small potential loss from failure to perform the same maintenance procedure on an old car with little expected lifetime capacity left.

Obviously, the subject of maintenance is a body of theory all its own which we do not treat in detail here.

Inventory Costs

We earlier identified the distinguishing characteristic of a durable as its existence beyond a single period. Because it has a life beyond a single period, it generates benefits and costs in common with all inventories of assets. We identify two inventory costs common to all inventories of assets as: time depreciation costs and control costs.

Time Depreciation Costs

Time depreciation cost is the difference between acquisition and salvage price in the period the durable is acquired and the change in the asset's salvage price in later periods as a result of factors other than changes in capacity. We have already discussed the possible costs associated with altering the physical capacities of the durable as a result of time, use, and maintenance. Now consider other factors which may alter an asset's salvage price. As a result of changing tastes and preferences, the demand for the output produced from the durable's services may change. New alternatives to the durable's services may be developed. The price and/or quality of companion inputs used with the durable may change. Inflation may change prices in general and the durable's in particular. Finally, the durable's value may change over time because the market in which the durable is traded is not perfect. If the market for the durable is perfect, the firm can buy and sell the durable at the same price. Most durable markets, though, are not perfect--try selling a new car for the same price as you paid the dealer--the car's price depends on whether you're buying or selling. All of these factors affect both the acquisition and salvage value of the durable, as well as the value in use of the durable.

But how do these external pricing considerations affect the firm's costs? The principle is: value the remaining lifetime capacity of the durable according to its opportunity cost. If the durable is owned by the firm, then it has two alternatives: to keep it in inventory or to sell it. If the firm keeps the durable, then one opportunity cost is the change in the salvage price of the durable between periods. This cost is referred to in this chapter as "time depreciation cost."

This method of costing the holding of inventories of assets raises a question of practical importance to utilities. For instance, should utilities charge for capacity held in inventory (and lost through use) according to the asset's book value or salvage price? While we recognize legal limitations in charging depreciation, for internal policies which determine the durable's use, cost should be accounted for in terms of the market or salvage price. We illustrate the issue involved with an example.

Suppose a restaurateur acquired a vintage wine 30 years ago for 50 cents. And suppose further that the restaurateur can sell the same wine today for $35 a bottle. A customer comes to the restaurant and orders the vintage wine with his meal. Should the restaurateur charge the customer 50 cents or $35 for the wine? Consider another example. Suppose a local shopkeeper acquired a large number of expensive spring-powered watches just prior to the marketing of a new, more reliable quartz watch. After the new quartz watches are marketed, the demand for the spring-powered watches drops off dramatically, lowering their market prices. Should the shopkeeper persist in his efforts to sell the spring-powered watches at their acquisition prices or at their lower salvage value?

The obvious answer is: that the restaurateur should charge $35 for his wine since the difference between 50 cents and $35 is the reward for his willingness to bear the risk of price changes, storage costs, etc., associated with the wine. In essence, the difference is a negative time depreciation cost. In the same vein, the shopkeeper must also sell at the salvage or market price, and the difference, a time depreciation cost, is again the (negative) reward for bearing risk. The salvage value is, afterall, merely the current market price, what one will pay today for the same item.

Perhaps the issue for utilities is not so clear as the two examples cited might lead us to believe. Suppose the time depreciation cost is negative, the prices of owned durables increase. Should utilities be allowed to charge off a higher user cost to customers because the price of the equipment has increased--charge $35 for the wine? If so, then shouldn't their returns include the capital gains? These are not easily settled questions.

One reason in favor of accounting for the loss of the durable's value in current dollars is to provide a fund for its replacement. If prices are rising and the user cost of the durable is based on book value rather than market value (in essence, ignoring time depreciation costs or benefits), then there will always be an inadequate accumulation of funds from direct user costs to finance the replacement. If inflationary pressures are responsible for increasing the durable's price, letting the price be the market price (which is constantly adjusted by the consideration of time depreciation costs) merely allows the owner of the durable to value the costs of using it up in constant real dollars. To charge customers for the use of the durable's services based on the depreciated book value of the durable when all prices have increased is, in effect, to subsidize their use of the durable's services.

In response to the question, "Should the negative time depreciation cost (a capital gain) be counted as an increased return?" the answer is yes, because, at the same time inflation is creating capital gains, it is increasing control costs which offset the capital gains.

The advantage of charging off user cost in terms of replacement cost is that it allows decision makers to maintain in proper perspective the relative value of inputs into the production process and to regulate their use accordingly. Here again, the issue of time is critical. If inputs are purchased and used up at the beginning of each period, the analysis will be always in terms of replacement cost--since replacement cost is merely the existing market price to the firm of its asset.

Control Costs

Holding inventories of assets entails costs other than simply the changing price of the durable. To hold an inventory of assets commits resources to those assets. It means those funds used to control the resource are not available for investment elsewhere. If equity funds are involved, the cost is the foregone earnings on the next best investment opportunity. If borrowed funds are involved, the cost is the interest paid on the loan and the cost associated with a reduced credit reserve. Whether equity or debt capital is involved, the cost of controlling assets in inventory is referred to as "control costs." The control costs will always be a function of inflation, since it accounts for changes in the buying power of resources committed to the control of the durable between periods. So, as prices rise and capital gains result, so will the control cost.

To illustrate control costs more explicitly, consider an asset A held in inventory for one period. Moreover, assume that there is no change in the asset's value during the period. If investment opportunities earning a rate of return r were available, but if, instead of earning return rA, the asset is held in inventory, then the value of those returns foregone, discounted to the present, equals $rA/(1+r)$--the control cost of holding asset A in inventory for one time period.

Indirect Capacity Costs

Indirect User Costs

Still another category of cost is that cost which measures the impact of current decisions to extract services from the durable on future control and time depreciation costs. Since control and time depreciation costs depend on the inventory of lifetime capacity held, decisions to use up capacity in the current period simultaneously affect time depreciation and control costs in the future. This opportunity cost is referred to as "indirect user cost."

Replacement Opportunity Cost

Current period use decisions do more than alter time depreciation and control costs in the future. They may alter the time when the durable is replaced or abandoned. To measure this cost consideration, we introduce our final cost--replacement opportunity cost. This cost equals the opportunities lost by failure to replace. Another example may help to illustrate this cost. Suppose a decision maker continues to drive an older car that gets poor gas mileage. Given that a new more fuel-efficient one is available, he entails an opportunity cost. The cost is equal to the fuel savings available from the new car. Of course, continuing to drive his older fuel-inefficient car, he incurs lower control cost which may affect the "replacement opportunity cost." But both are relevant in arriving at decisions of when to invest, disinvest, and use durable services. We will have more to say about replacement opportunity costs later on in this chapter.

Fixed versus Variable Costs

Often economists distinguish between costs that vary with use and costs that do not. The former are referred to as variable and the latter as fixed. In economic analysis in which a single period is considered, the time-related costs of control cost, capacity time cost, replacement opportunity cost, and time depreciation cost are considered fixed. Direct user cost and indirect user cost are not, varying as they do with output and use of the durable. Of course, in a single-period analysis, only direct user cost would be considered variable.

However, in single-period analysis in which the value of services from the durable exceeds its variable cost, the only relevant decision is the level of services to extract. The decision to disinvest/invest can only be considered when time is introduced and the expected time indexed benefits and costs are accounted for. From that long-run perspective, it is important to realize the importance of the costs we have identified for use and investment/disinvestment decisions.

The variable costs are critical for determining the level of services to extract from a durable, given that it exists in the firm. It should be clear that costs, opportunity costs, may be distinct from costs an accountant would compute. For example, indirect control costs or replacement opportunity costs would not appear in the profit and loss statement, which accounts for durable benefits and costs in a single time period. Nevertheless, their consideration is relevant for making use and investment/disinvestment decisions.

The distinction between fixed costs and variable costs is of some importance for utilities. Nuclear generators have large fixed costs, while user costs and other costs that vary with output are smaller in comparison. As a result, services should be extracted at near maximum operating capacity, since variable costs of increasing the service extraction rate are small relative to fixed costs. Base operating plants should be those with high fixed costs, such as nuclear plants, while those with high variable or use costs relative to time costs should be used under peak demand conditions. Gas and coal-fired power plants have variable costs which are more important relative to fixed costs. These are often older units, which are often largely depreciated so that further time depreciation and control costs are small. Replacement opportunity costs may be large, however, but these costs have more relevance to the replacement/investment decision, as the next section demonstrates.

Other Cost Considerations

Variable Costs and Durable Efficiency

So far in this chapter, we have had little to say about the relationship between durable and non-durable inputs, especially non-durable inputs which are divisible in use and acquisition. This relationship, we hypothesize, is critical in the investment/disinvestment use and maintenance decisions of the firm. Consider the parallel relationship between losses in lifetime capacity associated with use and losses in quantities of non-durable inputs used up in the production process. In both cases, the rates at which variable inputs are used and lifetime capacity is lost may vary with the rate at which services are extracted from the durable. We return to our passenger car to illustrate the point. At a service extraction rate of 55 miles per hour, our car may use up gasoline at the rate of slightly over 2.5 gallons per hour or slightly more than 22 miles per gallon. If the same car has services extracted at the rate of 75 miles per hour, the rate per mile of gasoline use might increase to 5 gallons per hour or 15 miles per gallon.

The relationship between rates of variable input use and service extraction rates can be illustrated graphically. In Figures 4.4a and 4.4b, the vertical axes measure the rates at which non-durable inputs are used during the time interval specified by the operating capacity measure. The horizontal axes, as before, measure the rate at which services are extracted from the durable. For example, if the operating capacity of a car is measured in terms of miles driven per hour and the variable input used is gas, then the vertical axis measures the gallons used in each hour as a function of the rate at which services are extracted from the durable.

Three possible relationships between variable inputs and durable extraction rates are illustrated in panels a and b of Figure 4.4. OA in panel a represents a flexible relationship, which gives rise to a constant average use of variable inputs A'A in panel b since the marginal use of inputs is a constant per unit increase in the rate services are extracted from the durable. A durable with increasing marginal use of variable inputs is depicted by curve OB in panel a and an increasing average use curve B'B in panel b. Note that, for this durable, the most efficient service extraction rate tends toward zero. Finally, curve OC in panel a represents a durable with marginal variable use which first decreases, then increases. This gives rise to an average use curve in panel b that first decreases, then increases, with its most efficient rate achieved at service extraction rate OD.

The transformation from average use of variable inputs to average variable costs can be made by multiplying the rate of inputs used by their acquisition price.

Other Considerations in Determining Service Extraction Rates

So far, this report has focused on how service extraction rates affect lifetime capacity of durables. Further, the report has emphasized the important relationships between the rate of service extraction, losses in lifetime capacity, and use of variable inputs. Costs associated with time are also important considerations, as they modify the optimal service extraction rates that would prevail without them. But other factors as well may be critical in determining optimal service extraction rates and optimal investments/disinvestments; that is, the costs of varying the rates of service extraction.

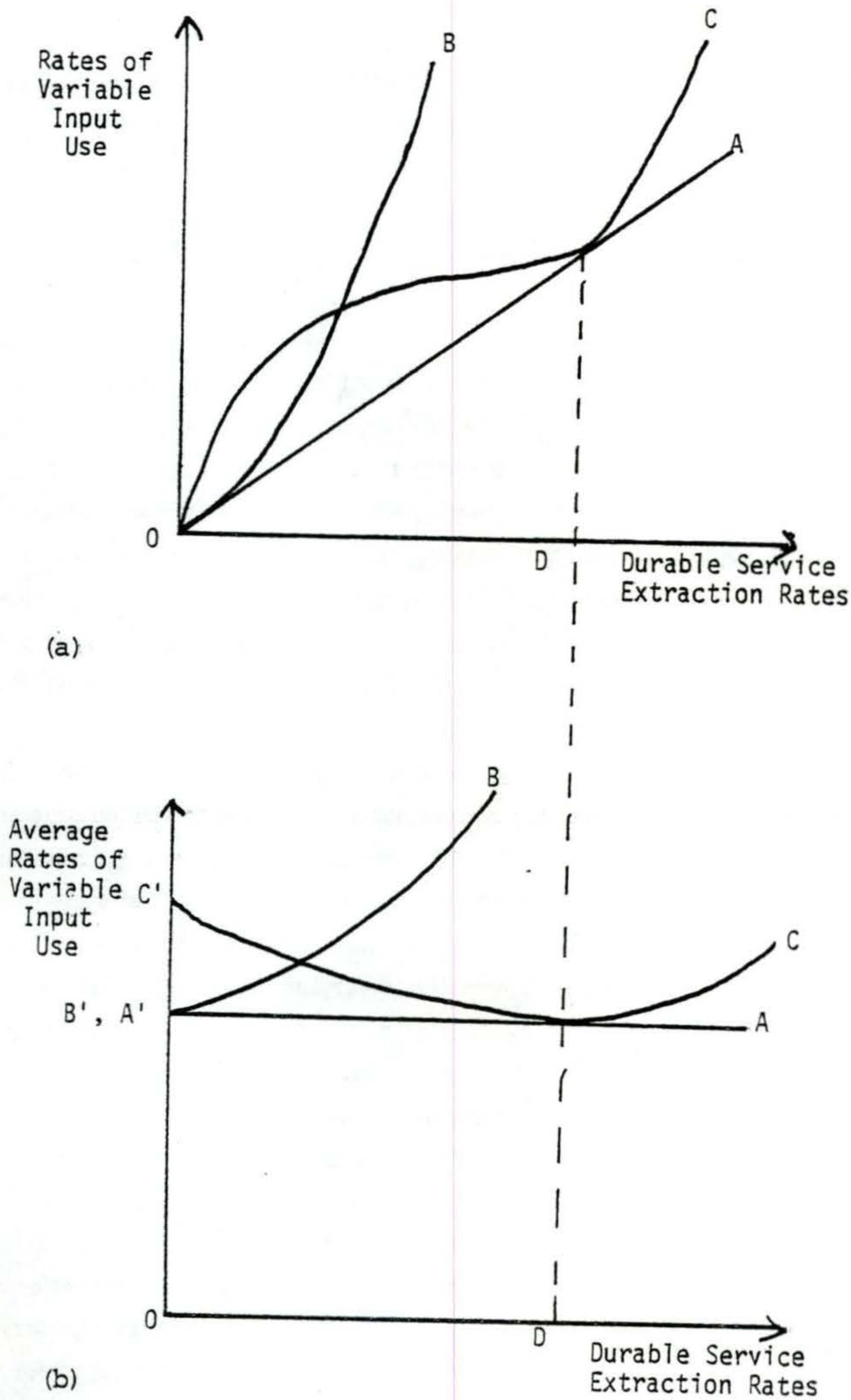


Figure 4.4

Relationship Between the Total and Average Amounts of Variable Inputs Used and the Amount of Services Extracted From a Durable

Particularly for utilities which face time-varying demand for their product, the ability and cost of modifying service extraction rates are critical features in the determination of optimal durable size and use. A plant which has a very low average loss in lifetime capacity and variable input use, yet has large expenses associated with changing service extraction rates, may be less preferred than one which has a higher average cost but which can adjust easily to varying service extraction rates.

The critical element in determining the relative advantages or costs of flexibility in adjusting to varying rates versus flexibility in varying losses in lifetime capacity associated with use depends on the variability of demand. If demand is constant, then there is little concern about losses in lifetime capacity due to varying the utilization rate. To achieve the optimal benefits from units that have low average costs in the face of time varying demands, utilities organize production between base and peaking units.

Base plants have the desirable quality that, although time costs are high, average costs can be made low by utilizing services at near maximum rates. For these units, near constant quantities of services are extracted. As a result, even though varying service extraction rates may be costly, base load units are not forced to meet those expenses because the service extraction rate is constant.

Peaking units, meanwhile, are utilized at varying rates. They meet the difference between the near constant levels of services supplied by the base load unit and the actual demand level. Typically, these units have smaller time costs than base load units and higher variable input costs. For example, the variable cost of services extracted from coal-fired generators likely exceeds those of nuclear. But the output from coal-fired generators can be adjusted much more easily to varying levels of demand.

A three-dimensional graph may help illustrate the relationship between current and past use decisions on losses in lifetime capacity. In Figure 4.5, let the two base axes represent services extracted in periods t and t+1. The vertical axis, meanwhile, equals losses in lifetime capacity. Along the 45-degree diagonal, services extracted in the two periods are the same, resulting in zero adjustment costs. Elsewhere, rates differ, and the greater the distance from the diagonal the greater the differences in the rates of services extracted in the two periods--and, as a result, the greater the adjustment costs.

To this point, the costs and benefits of owning and extracting services from a durable have been described qualitatively. Now, we add concreteness to our discussion by introducing explicit functions to describe the costs. In the process, we explicitly identify the indirect user cost and replacement opportunity cost. All of the costs are important when making investment/disinvestment decisions, but the relationship between replacement opportunity cost and the economic life of a durable is most closely linked. But

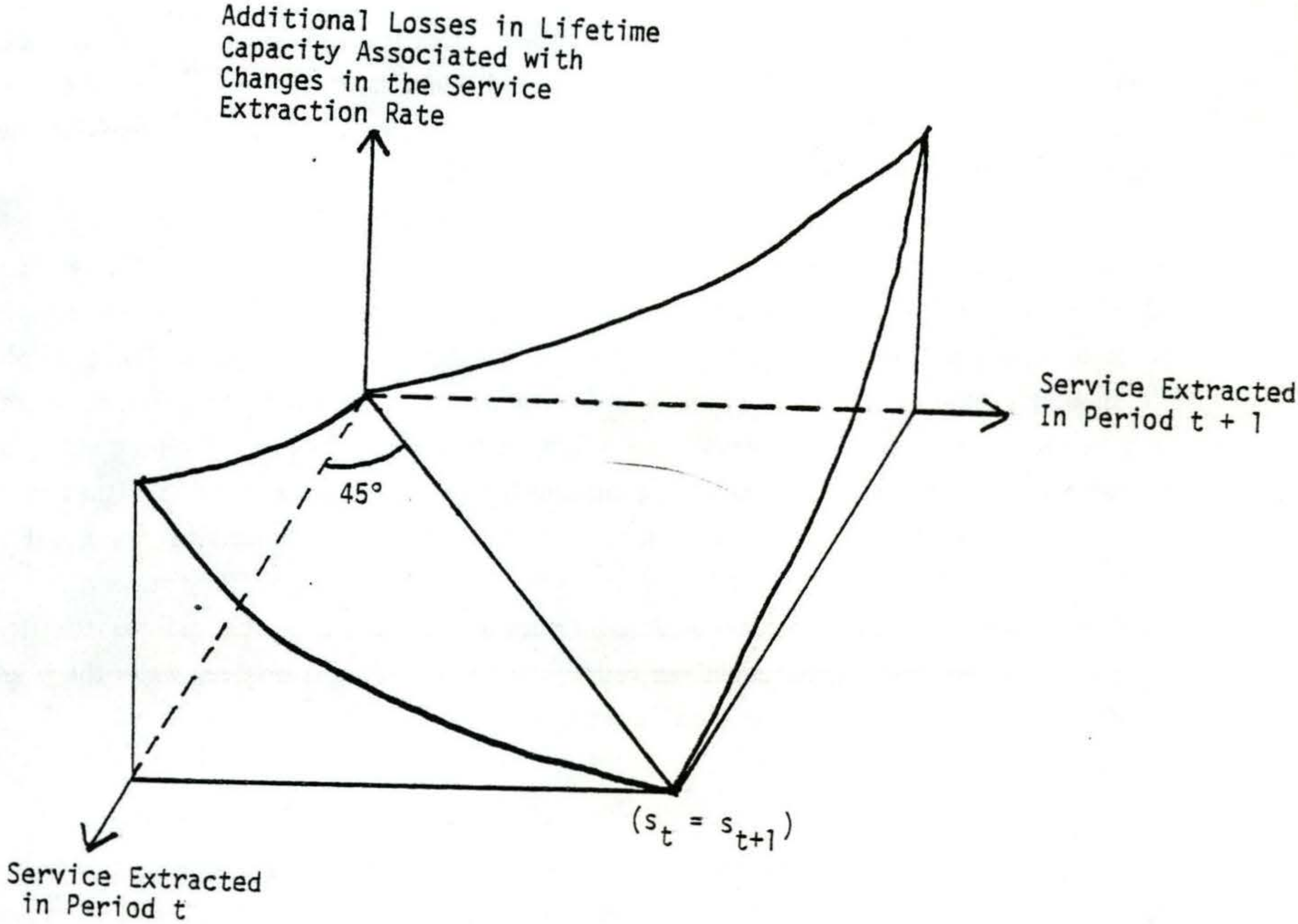


Figure 4.5

The Effect of Changing Service Extraction Rates on Losses in Lifetime Capacity before introducing explicit cost functions, we focus, for a moment, on benefits from durables.

Benefits from Durables

So far, the discussion of durables has been one-sided, focusing solely on costs. Obviously, at the time of acquisition, the expected benefits at least exceeded the cost. Completeness, then, requires that we at least survey some of the major benefits that might accrue as a result of durable ownership or control.

The principle reason for durable ownership is the acquisition of services to be used for producing goods of at least equal value to the cost of durable ownership. However, in addition to benefits accrued from the sale of a good produced using services from the durable, there are the benefits associated with ownership of all inventories. For example, the quality of the durable may improve over time, such as for wine. The market change may result in the price of the durable increasing as substitutes become more expensive or as demand for the good produced by the durable increases. Another benefit is that there may be a price advantage from buying in bulk. Instead of hiring a durable's service unit by unit, it may be cost advantageous to assume ownership of a large inventory of services. In addition to cost savings with associated bulk purchasing, buying in bulk assures the firm access to the services when needed--a reliability feature that is critical when there are seasonal or peak demands for the durable services.

Durable Decision Analysis

Identifying both benefits and costs of extracting services from a durable is the necessary step toward analysis of the several important decisions related to durables. These include: (1) the choice of the durable with appropriate capacities; (2) optimal time to disinvest/invest; (3) the optimal use of the durable; and (4) the optimal maintenance. For the moment, we will ignore the implications of uncertainty on these decisions.

All of these decisions are interdependent. The rate at which services are extracted influences the optimal life, the life of the durable influences the optimal capacity of the durable, etc. The interrelatedness of these decisions can be best demonstrated by solving a relatively simple replacement model of a production process which utilizes services from the durable and a variable input to produce an output. Moreover, to avoid, for the moment, the question of what size the durable should be, assume it initially has lifetime capacity F^*. Thus, because of lack of perfection in the durable asset market, let the average price per unit of remaining capacity at time t be $P(t)$, equal to:

$$(4.6a) \qquad P(t) = \bar{P}(1 + d)^{-t}$$

where $\bar{P}$, as defined earlier [equation (4.3)], equals the average acquisition (salvage) price per unit of lifetime capacity. The change in average price from one time period to the next equals:

$$(4.6b) \qquad P(t) - P(t-1) = - d\,\bar{P}\,(1 + d)^{-t}$$

which, when multiplied by the current capacity of the durable F(t), equals time depreciation cost TD(t):

$$(4.7) \qquad TD(t) = d\,\bar{P}\,(1 + d)^{-t}F(t)$$

Next, we identify control costs associated with the durable. This cost at time t depends on the durable's remaining lifetime capacity F(t), the rate of return r that could be expected in the next best investment opportunity, and the average price of the durable's remaining capacity. Identifying the control cost for the t-th period as CC(t), we can write:

$$(4.8) \qquad CC(t) = r\,\bar{P}\,(1 + d)^{-t}\,F\,(t)$$

To add concreteness to the analysis, while maintaining an essential amount of tractibility, let the durable used in the productive process be less than perfectly flexible and express the relationship at the beginning of the t-th period between remaining lifetime capacity F(t), initial capacity F*, and services extracted from the durable Z(t) as:

$$(4.9) \qquad F(t) = F^* - \sum_{j=1}^{t-1} (Z^2(j)\,b + a)$$

In equation (4.9), b and a are parameters, ta is the accumulated loss in capacity due to time, while $\Sigma\, Z^2(j)\, b$ is capacity lost due to use. The loss of lifetime capacity in the t-th time period is:

$$(4.10) \qquad F(t+1) - F(t) = - [Z^2(t)\,b + a]$$

while the marginal loss in capacity due to increased use is:

$$(4.11) \qquad dF(t)\,/\,dZ(t) = -2\,Z(t)\,b$$

Dividing (4.10) by Z(t) gives the average loss in lifetime capacity per period, which equals the marginal loss in equation (4.11) when Z(t) equals the rated capacity. Solving for the rated capacity by setting the average loss equal to marginal loss, we obtain:

$$(4.12) \qquad Z(t) = (a/b)^{1/2}$$

If Z services are being extracted at the rated capacity, then u(t), the utilization of the durable, equals one. By varying the utilization rate, services extracted are also varied, a relationship that can be expressed as:

$$(4.13) \qquad Z(t) = u(t)\,(a/b)^{1/2}$$

Because service extraction rates are always linked to some time period (miles per hour, e.g.), it will be convenient to discuss service extraction rates in terms of the utilization rate (a percent). This switch will focus attention on the percentage of rated capacity being used rather than time.

To express lifetime capacity in terms of utilization rates, we substitute the right-hand side of (4.13) into equation (4.9). Then, the remaining lifetime capacity at the beginning of period t can be written as:

$$(4.14) \qquad F(t) = F^* - \sum_{j=1}^{t-1} [1+u^2(j)]\,a$$

while losses in lifetime capacity during the t-th period equal:

$$(4.15) \qquad F(t+1) - F(t) = -\,[1+u^2(t)]\,a$$

Finally, a more explicit expression for lifetime capacity can be determined once the calendar life of the asset is set. Suppose the durable will last t* periods if operated at its rated capacity. Summing (4.15) for t* periods with u(t) equal to 1, F* can be written as:

$$(4.16) \qquad F^* = t^*\,2a$$

Having defined the relationship between loss in capacity, time, and use, direct user cost (DUC) and capacity time cost (CTC) become simply the loss in capacity per unit of time multiplied by the average market price of the lost capacity, or:

$$(4.17) \qquad CTC(t) + DUC(t) = a\,[1+u^2(t)]\;\bar{P}\,(1+d)^{-t}$$

If a variable input x(t) in the t-th period is also required in the production of services from the durable, we include still another cost, the cost of the variable input [VC(t)] equal to:

$$(4.18) \qquad VC(t) = p_x\,x(t)$$

where p_x is the price per unit of x(t). Let the requirement of x(t) per unit of durable service used increase in an inflexible relationship with the amount of services extracted, such as:

$$(4.19) \qquad x(t) = c\,u^2(t)\,(a/b)$$

Equation (4.19) now allows us to express our variable costs, in terms of durable services extracted, as:

$$(4.20) \qquad VC(t) = p_x \, cu(t)^2 \, (a/b)$$

Having identified explicitly our cost functions, we combine direct user cost, capacity time cost, variable cost, time depreciation cost, and control cost into a single total cost function [TC(t)] and write:

$$(4.21) \qquad TC(t) = [(1+u^2(t))a + (r+d)\,F(t)]\ \bar{P}(1+d)^{-t} + p_x\, cu^2(t)\,(a/b)$$

subject to: $F(t) = t^{*}2a - \sum_{j=1}^{t-1} a\,[1+u^2(t)]$ and $F(t) \geq 0$

Then, if services of the durable are valued at a constant price p, gains g(t) in the t-th period can be written as:

$$(4.22) \qquad g(t) = p\, u(t)\, (a/b)^{1/2} - TC(t)$$

Suppose we applied our static analysis techniques to (4.22) and differentiated with respect to u(t) to find the optimal service extraction level. The result would be a service extraction rate which equates current period returns to current period costs.

$$(4.23) \qquad \partial g(t) / \partial u(t) = 0$$

This approach would, however, ignore a critical cost element, an indirect capacity cost, which this chapter identifies as indirect user cost.

Introducing Indirect User Cost

In a static (timeless) analysis, the cost of using up lifetime capacity is the loss in the durable's salvage value because of use--direct user cost. But there is another cost (benefit) that enters when the analysis is dynamic. It is the effect on future control costs resulting from changes in the durable's lifetime capacity through use in the current period.

All other things being equal, increasing the use of the durable in the current period reduces its lifetime capacity, which in effect reduces available durable services held in inventory. This reduction in turn reduces future control costs. The present value of those saved control costs is part of indirect user costs.

There is another part, however. Suppose that the price of the durable is increasing (decreasing), so that on the remaining lifetime capacity of the durable, there is a capital gain (loss) realized. Then, to use up the durable in the current period implies a reduction in future capital gains (losses), as well as the requirement to eventually replace with a higher (lower) priced durable.

To account for these influences requires that we write a multi-period gain function which includes benefits and costs of extracting services from a durable over its lifetime. Forming the multi-period gain function G over the economic life s of the durable, we write:

$$(4.24)\quad G = p\,u(1)\,(a/b)^{1/2}\,(1+r)^{-1} + \ldots + p\,u(s)\,(a/b)^{1/2}\,(1+r)^{-s}$$
$$- \{[(1+u^2(1))a + (r+d+rd)\,F^*]\quad (1+d)^{-1}\,\overline{P} + p_x\,u^2(1)\,c\,(a/b)^{1/2}\}\,(1+r)^{-1}$$
$$- \ldots$$
$$- \{[(1+u^2(s))a + (r+d+rd)\,F(s)]\quad (1+d)^{-s}\,\overline{P} + p_x\,u^2(s)\,c\,(a/b)^{1/2}\}(1+r)^{-s}$$

such that

$$F(t) = t^*\,2a - \sum_{j=1}^{t-1}\,[1+u^2(j)]\,a$$

and

$$F(t) \geq 0 \qquad \text{for } t=1, \ldots, s$$

In the first line of (4.24), services extracted in each of the j periods, $u(j)\,(a/b)^{1/2}$ ($j = 1, \ldots, s$), are multiplied by their output price p to obtain total returns. Then, returns are discounted to the present value by the term $(1+r)^{-j}$ so that comparisons are made in present value dollars. On the following lines, direct user costs and capacity time costs, control and time depreciation costs, and variable costs, are subtracted and discounted to the present.

Consider now the optimal utilization rate in the j-th period. Gross returns in the j-th period obviously depend on u(j), but so do control and time depreciation costs in all later periods. That is, utilization of the durable in the j-th period affects the lifetime capacity of the durable in all future periods. The lifetime capacity in future periods affects control and time depreciation costs. Therefore, future time depreciation and control costs will affect optimal utilization rates in the current period.

To illustrate this simultaneity between utilization rates and future period control and time depreciation costs, we differentiate (4.24) with respect to u(j), which results in:

$$(4.25)\quad \partial G / \partial u(j) = p\,(a/b)^{1/2} - 2\,u(j)\,a\,(1+d)^{-j}\,\overline{P} - 2\,u(j)\,p_x\,c\,(a/b)^{1/2}$$
$$- (1+r)^j\,(r+d+rd)\,\overline{P}\,\sum_{i=j+1}^{s}\,[\partial F(i) / \partial u(j)]\,(1+d)^{-i}\,(1+r)^{-i}$$

The right-hand side of (4.25) gives the returns per unit change in utilization $[p\,(a/b)^{1/2}]$ and subtracts the increases attributable to indirect user cost, increases in the variable input, and, in the last term, indirect user cost--that is, the increases in time depreciation cost and control cost in future periods resulting from the use decision in period j.

Since $\partial F(i) / \partial u(j)$ for all values of i equals a negative 2a u(j), indirect user cost has a sign effect opposite to direct and variable input costs. Making the substitution for $\partial F(i) / \partial u(j)$ and summing the geometrically weighted indirect user cost, we can solve for u(j) in (4.25) as:

$$(4.26) \qquad u(j) = p / 2 \, [p_x c + (ab)^{1/2} (1+d)^{-s} (1+r)^{-s+j} \bar{P}]$$

The importance of indirect user cost in determining u(j) in (4.26) depends critically on three elements of the multi-period gain function: (1) the remaining life of the durable; (2) the rate of change in the average price of the durable's remaining capacity; and (3) the discount rate used to convert future benefits and costs to their present value equivalents. All of these influences will play an important role in determining the supply of the durable's services.

Supply of Service Functions

The analysis thus far has assumed the firm operates in a perfectly competitive market in which it receives the market price p and could supply as much as it wished without affecting the market price. A justification for these stringent assumptions might be that, in a partial analysis conducted for a large utility, a single plant may produce to its capacity, receive the regulated price, and not have its regulated price affected.

In any event, retaining these assumptions allows us to consider the firm's supply response of services. Obviously, the supply will change in each period, requiring, therefore, a period by period analysis.

The general form of the supply response function given in (4.26) is linear in the output price p, as Figure 4.6 illustrates. The slope of the function depends on the remaining life of the durable, variable input price, discount and time depreciation rates, average acquisition price of the durable, and parameters of the variable input use and direct user cost functions. As one would expect, a decrease in the average acquisition price $\bar{P}$ in period t+1 rotates the supply function upward from OA to OB in Figure 4.6 as the cost of using up the durable's lifetime capacity is reduced. A similar response is obtained by decreasing the variable input price p_x, which decreases the cost of using up the variable input.

Consider also the effect on the supply of services from the durable as the durable's remaining life s is increased. As the remaining life is increased, control costs and time depreciation costs become more important and any reduction by use in the inventory of lifetime capacity in near periods will be even more beneficial with a longer lifetime. To demonstrate this result using equation (4.26), we find the new service level u(j) as s becomes large. The result is the limit:

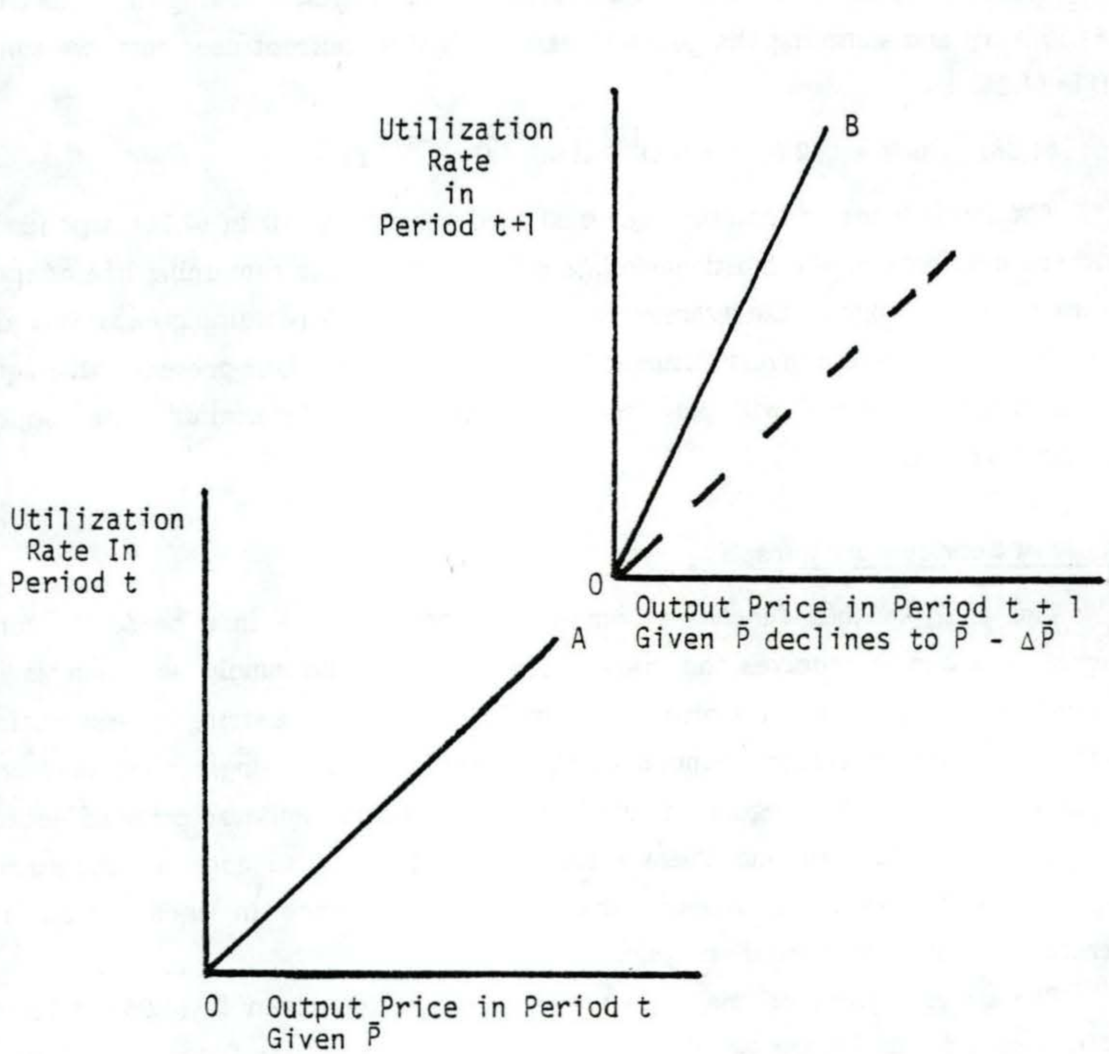


Figure 4.6

A Supply-of-Service Response to Increases in Output Price P and the Durable's Average Capacity Price $\bar{P}$

$$(4.27) \quad \lim_{s \to \infty} u(j) = p/2p_x c$$

which is greater than u(j) in (4.26) for a finite s. That is, making s larger increases the utilization rate u(j) for a given output price. The result, then, would be a supply shift from OB to OA.

Next, consider the effect of increasing the amount by which the average price of the durable's lifetime capacity declines in each period--increasing d. In this case, increasing the amount by which the average price of the durable's capacity declines in each period increases time depreciation costs and reduces direct user costs. This effect alone increases the utilization rate. In addition, increasing d means that greater time costs will be incurred in the future on remaining capacity, a loss which can be reduced by increased use of the durable in current periods. As a result, increasing d increases utilization of the durable:

$$(4.28) \quad d\,u(j) / dd = 2\,p(ab)^{1/2}\, s\,(1+d)^{-s-1}\,(1+r)^{-s+j}\,\bar{P} / 4\,[\,.\,]^2 > 0$$

where the bracketed dot equals the denominator in (4.26). Graphically, this represents an upward rotation of the supply of services from OB to OA in Figure 4.6.

Finally, increasing the discount rate r has the same effect as increasing d--it makes the cost of carrying capacities or inventories into the future more expensive. As a result, increasing r increases the rate services are extracted, thereby reducing future inventories:

$$(4.29) \quad d\,u(j) / dr = 2\,p(ab)^{1/2}\,(s-j)\,(1+r)^{-s+j-1}\,(1+d)^{-s}\,\bar{P} / 4\,[\,.\,]^2 > 0$$

again producing an upward rotation in the durable's supply-of-services function in Figure 4.6.

Finally, it may be of some interest to determine the pattern of usage the durable will likely experience over time. That is, how will the optimal u(j) compare with the optimal u(j+1)? The relationship can be found simply by updating the time subscript in (4.26) by one period and by forming the ratio between u(j) and u(j+1). The result is:

$$(4.30) \quad u(j+1) / u(j) = [p_x c + (ab)^{1/2}\,(1+d)^{-s}\,(1+r)^{-s+j}\,\bar{P}] \;/\; [p_x c + (ab)^{1/2}\,(1+d)^{-s}\,(1+r)^{-s+j+1}] < 1$$

That is, time has reduced the importance of time and control costs, thereby reducing the incentive to use up resources now. Consequently, service extraction levels can be expected to decline over time even if the durable's price is increasing.

Including Replacement Opportunity (RO) Cost

Our earlier analysis described costs which help determine the optimal level of services to extract from a durable. One problem, however, of formulating at least one of those costs was that we assumed that the remaining life of the durable was known. Obviously, some limits on the durable's life can be imposed--namely, it can't have a life beyond which its available services are exhausted, and it won't have an economic life beyond the period in which returns are less than costs. Apart from these upper limits, however, we have not determined how long a durable should be kept, especially if the durable will be replaced by another.

To find the durable's optimal life, we begin by expressing the net gains in the t-th period as g(t), equal to the difference between gross returns, direct user cost, capacity time costs, plus time depreciation and control cost in the t-th period. This, of course, assumes that optimal utilization rates u(t) have been determined, a problem we ignore for the moment. Then, the integral of g(t) discounted by the discount rate r equals our multi-period gain function G, or:

$$(4.31) \qquad G = \int_{o}^{s} g(t)\, e^{-rt}\, dt$$

The multi-period gain function above is expressed in continuous form to allow us to differentiate with respect to time. Assuming that g(t) is optimal, i.e., that optimal service extraction rates have been determined, we maximize G with respect to s and obtain:

$$(4.32) \qquad g(s) = 0$$

That is, produce until the revenues in the last period equal the sum of user cost, time cost, and control cost. The control cost in the last period is merely the rate of return r obtainable in the next best use times the salvage value of the durable, while the time and user cost equal the actual decline in the value of the durable associated with time and use.

Suppose, however, there exists a replacement durable, at least as efficient as the one in use. To continue with the older, less efficient durable, then, entails a cost, a cost equal to the net returns lost by not replacing the durable.

In order for there to be a replacement opportunity cost in our model, there must be at least one replacement considered. Assume we begin our planning horizon with durable F_1 which generates optimal net returns g(t) for s_1 periods. Then, at the end of the s_1 periods, F_1 is replaced with durable F_2, which generates optimal net returns $g_2(t)$ for s_2 periods. Under these assumptions, we reformulate our G function as G*:

$$(4.33) \qquad G^* = \int_0^{s_1} g_1(t)\, e^{-rt}\, dt + e^{-rs_1} \int_0^{s_2} g_2(t)\, e^{-rt}\, dt$$

and ask: is s_1, the economic life of F_1 with replacement F_2, the same as s, the optimal economic life for F_1 without a replacement?

The first order conditions for G* with respect to s_1 can be written as:

$$(4.34) \qquad g(s_1) = r \int_0^{s_2} g_2(t)\, e^{-rt}\, dt$$

We can infer, since $g(s_1)$ in (4.34) does not equal g(s) in (4.32), that s_1 and s cannot be equal. In fact, when a replacement is considered, we no longer produce until net returns equal zero, as (4.32) suggests, but rather until the returns in the last period equal the returns foregone by postponing the replacement of the durable by one period. That postponement cost we have already defined as the replacement opportunity (RO) cost, which equals the right-hand side of (4.34).

The presence of an RO cost with net returns from the durable diminishing over time should result in a shorter economic life for the durable in use. In fact, the more replacements considered, the shorter the life of the existing durable, as we will shortly demonstrate. We state this result only tentatively, however, since altering the durable's life alters indirect user cost and optimal service extraction rates. Our earlier results [see equation (4.27)] suggest that utilization should decrease, as shortening the time reduces the advantage of reduced inventory from increased use. How lower utilization rates influence the time pattern of g(t) remains unexamined.

Consider the complication introduced by the RO cost. To find s_1 in (4.34) requires us to know s_2. We could, of course, optimize (4.34) with respect to s_2, but in the process we would be forced to assume F_2 has no replacement, i.e., no RO cost. Another alternative is to assume that the economic history of the second durable duplicates the first, in which case the function G can be written as:

$$(4.35) \qquad G = \int_0^{s_1} g(s)\, e^{-rt}\, dt + e^{-rs_1} \int_0^{s_1} g(s)\, e^{-rt}\, dt$$

Now, the first-order condition can be written as:

$$(4.36) \qquad g(s_1) = r\,(1 + e^{-rs_1})^{-1} \int_0^{s_1} g(s)\, e^{-rt}\, dt$$

Alternatively, we could assume, as Perrin (1972) has done, an infinite number of identical replacements for F*, in which case the first-order condition for determining the economic life of the first durable is:

$$(4.37) \qquad g(s_1) = r\,(1 - e^{-rs_1})^{-1} \int_0^{s_1} g(t)\, e^{-rt}\, dt$$

Comparing (4.36) and (4.37), some important observations can be made. First, as long as the function g(t) is concave, adding the RO cost to time, control, and user costs shortens the optimal s_1. Secondly, increasing time and the number of durable replacements increases the RO cost and as a result shortens s_1 even more.

In addition, Perrin (1972) observed that in (4.37) the sign of ds_1/dr could not be determined. This result occurs because increasing r increases the opportunity costs associated with postponed consumption while at the same time reduces the value of postponed consumption.

To summarize, the inclusion of RO cost has created somewhat of a dilemma. If we assume that replacements are not identical, then we cannot find the optimal life of the first durable because it depends on the optimal life of the second one, but the life of the second one depends on the life of the third one, and so on.

The alternative is to assume in the two-asset replacement model that the second durable is identical to the first. The resulting first-order conditions, then, do not include the interdependency between the life of the current and successive durables. Once we assume the second durable is like the first, however, why not assume the same for the third, fourth, and so on? The result is (4.37), which is greater, as expected, than (4.36) which considers only one durable replacement.

Relaxing An Assumption

Once we recognize the stringent nature of the assumptions required to obtain a deterministic solution for s_1, an effort should be made to relax, if possible, some of the assumptions. This section relaxes one of them: it introduces a multiplicative technological change at the beginning of each durable's life.

Assume that technological advances increase the returns and costs in each period of the first durable's life by a factor (1+i), for the second durable by $(1+i)^2$, and so on. How would such a development affect the length of the durable's economic life? Under such an assumption, G can be expressed as:

$$(4.38) \qquad G = \int_0^{s_1} g(t)\, e^{-rt}\, dt + (1+i)\, e^{-rs_1} \int_0^{s_1} g(t)\, e^{-rt}\, dt + (1+i)^2\, e^{-2rs_1} \int_0^{s_1} g(t)\, e^{-rt}\, dt + \ldots$$

Summing the above expression geometrically, and if $(1+i)\, e^{-nrs_1}$ approaches zero as n becomes large, we can write G as:

$$(4.39) \qquad G = [1 - (1+i)\, e^{-rs_1}]^{-1} \int_0^{s_1} g(t)\, e^{-rt}\, dt$$

and the first-order conditions, found as before by differentiating (4.34) with respect to s_1, are:

$$(4.40) \qquad g(s_1) = r\, [(1+i)^{-1} - e^{-rs_1}] \int_0^{s_1} g(t)\, e^{-rt}\, dt$$

Except for the presence of i in the denominator, (4.40) is identical to (4.37). As a result, introducing i increases RO costs and shortens s_1. Intuitively, it follows that, if succeeding durables are technologically improved, delaying their installment will be more costly. Our intuition can be confirmed mathematically by differentiating (4.40) totally, first with respect to s_1 and then with respect to i. The result is:

$$(4.41) \qquad ds_1 / di < 0$$

The replacement criterion just deduced depends on the second-order condition that the gain function over time is concave. To assume otherwise produces some variations in the results, which are now considered.

Suppose the net gain (g) in each period is constant. Then, the discounted sum of the gain function could be written as:

$$(4.42) \qquad G = g\,[1 - (1+r)^{-t}] \;/\; r$$

Meanwhile, the annualized value of G is equal to:

$$(4.43) \qquad g = r\,G \;/\; [1 - (1+r)^{-t}]$$

Our replacement criterion says to replace the durable when the last period's return (in this case, g) is less than the annualized gain function G (in this case, also g). As a result, the stopping criterion is never met. This explains why some assets such as land, with near constant real returns, are seldom sold by the owner until his retirement or death.

Now, consider two other alternatives: (1) that g increases monotonically over time, and (2) that g decreases monotonically over time. Examining the first case, let $g(t) = g\,(1+w)^t$, where w is positive. Then, consider the gain function G with g(t) increasing, which can be written as:

$$(4.44) \qquad G = g\,(1+w)\,[\,1 - (1+w)^t\,(1+r)^{-t}\,] \;/\; (r-w)$$

Next, we annualize G by multiplying both sides of (4.44) by the annuity factor $r / [\,1 - (1+r)^{-t}\,]$. Then, we compare this annuity value in the t-th period to the t-th period gain function value, equal to $(1+w)^t\,g$.

$$(4.45) \qquad gr\,(1+w)\,[\,1 - (1+w)^t\,(1+r)^{-t}\,] \;/\; (r-w)\,[\,1 - (1+r)^{-t}\,] \overset{?}{=} g(1+w)^t$$

Cancelling the g value and multiplying both sides by $(1+w)^{-t}$, we can show that the annuity value always exceeds g, and, as a result, the stopping criterion is never met and the durable is kept until lifetime services equal zero.

$$(4.46) \qquad r(1+w)\,[\,(1+w)^t - (1+r)^{-t}]\,/\,(r-w)\,[\,1 - (1+r)^t\,] > 1$$

Using a similar approach, we can show that, if g is monotonically decreasing, the durable should be replaced every period.

Having answered one more part of the durable investment/disinvestment problem, we are prepared to examine the holistic approach to solving the problem. This leads us to a consideration of the interdependencies between the different aspects of the problem.

An Iterative Approach for Resolving the Dependency Between Indirect User Cost and Replacement Opportunity Cost

An interesting dependency has been created in our analysis that appears to require an iterative approach to resolve. That is, to determine optimal service extraction levels, we are required to know indirect user costs. In order to determine indirect user cost, however, we are required to know the remaining life of the durable.

A solution to this interdependency problem is to begin by assuming the optimal life is known, i.e., choosing s. Then, solve for optimal utilization rates u(1), . . . , u(s) by differentiating the G function with respect to each. Finally, compare the returns in the s-th period g(s) with the annualized average $G\, r / [1 - (1+r)^{-s}]$. If g(s) equals or exceeds $G\, r / [1 - (1+r)^{-s}]$ while g (s+1) is less than $G\, r / [1 - (1+r)^{-s-1}]$, the optimal time period has been found. If the last period's return exceeds the annualized average of the multi-period gain function, the time period of analysis selected was too short and should be increased and the procedure repeated; if returns are less than the annualized average, the period of analysis was too long and should be shortened and the calculation repeated.

After completing the above analysis, it could be repeated for alternative durables and comparisons made on the basis of their annualized average return.

We summarize the procedure as follows:

STEP 1 -- Select the durable investment to be analyzed (which may be a durable in place). Also, choose a best guess of the durable's remaining economic life. Form the multi-period gain function--an example was given in equation (4.24). Designate the current period as period 1 and the last period as period s.

STEP 2 -- Find optimal values for u(1), . . ., u(s) and, if applicable, find optimal values for operating and maintenance inputs x_1 (t), x_2 (t), . . ., (t=1, . . ., s), by differentiating the multi-period gain function G with respect to u(t) and the other variable inputs, accounting for all direct and indirect capacity costs in the expressions for costs and returns. Represent these optimal values as $u^*(t)$, $x_1^*(t)$, $x_2^*(t)$, . . ., for t=1, . . ., s.

STEP 3 -- Identify the net returns in each period by subtracting from gross returns: direct user cost, capacity time costs, variable costs, time depreciation cost, and control cost. The remainder should equal returns attributed to the durable as long as returns result only from services generated by the durable.

STEP 4 -- Consider the expression below:

$$(4.47) \qquad g(s) > r\,[\,1 - (1+r)^{-s}]^{-1}\; G$$

If the expression is true, then the marginal contribution of the durable in the last period exceeds its annualized average of a replacement with an identical economic performance--so s should be increased. Return to STEP 1. If the expression

$$(4.48) \qquad g(s) < r\,[\,1 - (1+r)^{-s}\,]^{-1}\; G$$

is true, then s should be shortened. That is, the last period's net gain reduced the annualized average and a higher annualized average return could be realized by shortening the economic life of the durable. After choosing a shorter life, then return to STEP 1.

Only if:

$$(4.49) \qquad g(s) \geq r\,[\,1 - (1+r)^{-s}\,]^{-1}\; G \quad \text{and} \quad g(s+1) < r\,[1 - (1+r)^{-s-1}]^{-1}\; G$$

are true, does the analysis stop. The optimal life as well as the optimal service extraction rates are now determined; moreover, the expression $r\,[\,1 - (1+r)^{-s}\,]^{-1}\; G$ provides an index of the durable's performance--in essence, a time adjusted average return.

Lastly, if more than one durable is under consideration, then the procedure is repeated. The final result again is an annualized average return for each durable, which leads us to the last step.

STEP 5 -- Choose the durable with the largest annualized average return and acquire it if the net present value of G is positive.

Valuing the Durable's Services

Until now, we assumed that returns to the durable's services could be identified and measured independently of other inputs. It should be clear that the value of the durable's services is not so easily identified. For example, suppose more than one input was involved in the production process. Then, returns to the durable must be identified separately to determine its optimal economic life and to determine if the durable's returns covered all costs. But how are returns attributable to the durable identified? Historically, assigning the value of production to inputs has been of interest to economists. Early studies were interested in determining the share of a process labor could claim when it was combined with capital and other inputs. Farm management experts have long dealt with the issue of how to identify the returns to farmland so as to establish the maximum bid price that could be offered.

The two efforts to identify returns to factors of production resulted in two different methodological approaches. As a theoretical tool, economists applied Euler's theorem to separate the output among the inputs. A practical approach adopted by farm managers was the residual approach, which assigned to land what was left of returns after paying for variable expenses.

Neither Euler's theorem nor the residual approach ultimately solved the problem of how to assign the output to its inputs. Euler's theorem, which said each input's share of the output was its marginal value product of the last unit produced times the input level, was applicable only for linearly homogeneous production functions; and the residual farm management approach has no basis for assigning net surplus from production to land any more than they could defend assigning it to labor, machinery, or capital.

A New Theorem

Faced with an apparent impasse on how to value durable asset services, hence durables, this paper proves a theorem which states that, by specifying relationships between inputs, the output of any continuous production process can be divided among the inputs in such a way that the output is just exhausted.

Critical to the theoretical development is the understanding that, although in many cases durables are lumpy in acquisition, they provide services that are extracted in completely divisible amounts. Therefore, being compatible with traditional marginal economic analysis, we can ask: how many units of service should we extract from the durable? In contrast to marginal analysis, however, we still face the question: is the total value of services extracted from the durable at least equal to its capacity and inventory costs? The following theorem will allow us to answer such a question.

The Product Exhaustion Theorem

Theorem 1: Let $f(x,y)$ be a continuous function with derivatives $f_x(x,y)$ and $f_y(x,y)$. If the relationship between x and y is specified, say $y = m(x)$ or $x = m^{-1}(y)$ for all x and y, then:

$$\int_x f_x(x,m(x))dx + \int_y f_y(m^{-1}(y), y)dy = f(x,y)$$

Proof: After substituting $m(x)$ for y, express the function $f(x,y)$ as $f(x,m(x))$. The derivative of $f(x,m(x))$ with respect to x can be written as:

$$df = f_x(x,m(x))dx + f_y(x,m(x))m_x(x)dx$$

The anti-derivative of df, which by the second fundamental law of calculus equals f, can be written as:

$$f(x,m(x)) = \int_x f_x(x,m(x))dx + \int_{m(x)} f_y(x,m(x))m_x(x)dx$$

Then, by using the change-of-variable technique, we substitute dy for $m_x(x)dx$, and y for m(x) in the second integral to obtain the desired result. The change-of-variable technique then allows us to substitute in the above expression y for m(x); $m^{-1}(y)$ for x, and dy for $m_x(x)dx$. We write the result as:

$$f(x,y) = \int_x f_x(x,m(x))dx + \int_y f_y(m^{-1}(x),y)dx$$

which proves our theorem.

At this point, an interpretation of the product exhaustion theorem may prove helpful. The interpretation is aided by Figure 4.7, which illustrates isoquants q_1, q_2, . . ., etc. The isoquants represent constant levels of output obtained from various combinations of inputs x and y used in the process f(x,y). The isoquants are connected by ray OA, which describes an expansion path.

Now, consider output levels q_1 and q_2 obtained with inputs (x_1,y_1) and (x_2,y_2), respectively, at points B and C along the expansion path OA (Figure 4.7). The increase in output from q_1 to q_2, moving from point B to point C, can be achieved by increasing the input x by an amount (x_2-x_1) and by increasing input y by an amount y_2-y_1. The increase in output that results, q_2-q_1, is a result of increases in both x and y. Were only x increased, output would have only increased by q_o-q_1, the remaining increase, q_2-q_o, being attributed to the increase in y equal to y_2-y_1. Output q_o-q_1, then, approximates the contribution to output of the x input at output level q_1. A similar contribution could be obtained for increases in the input y by the amount y_2-y_1. These measures only approximate the contributions of x and y, however, since, along the expansion path, x and y are changing simultaneously. Output increases measured at smaller and smaller increases in x and y would, though, improve the accuracy of our measure. Moreover, for infinitely small changes in x and y, completely accurate measures of the output contributions of x and y are obtained and the results are the same as those obtained with the product exhaustion theorem.

Now, after having established a methodology for assigning the output to the inputs, we are ready to ask the next question: what is the proper relationship to define between the inputs? This question, of course, is an economic one which depends on the marginal value products of the inputs and their marginal costs. At a point in time, this relationship depends only on those costs that vary with use as opposed to costs that vary with time. Economists prescribe that the relationship among the inputs should be such that the ratios of their marginal value products divided by their prices be equal for all inputs at any point in time.

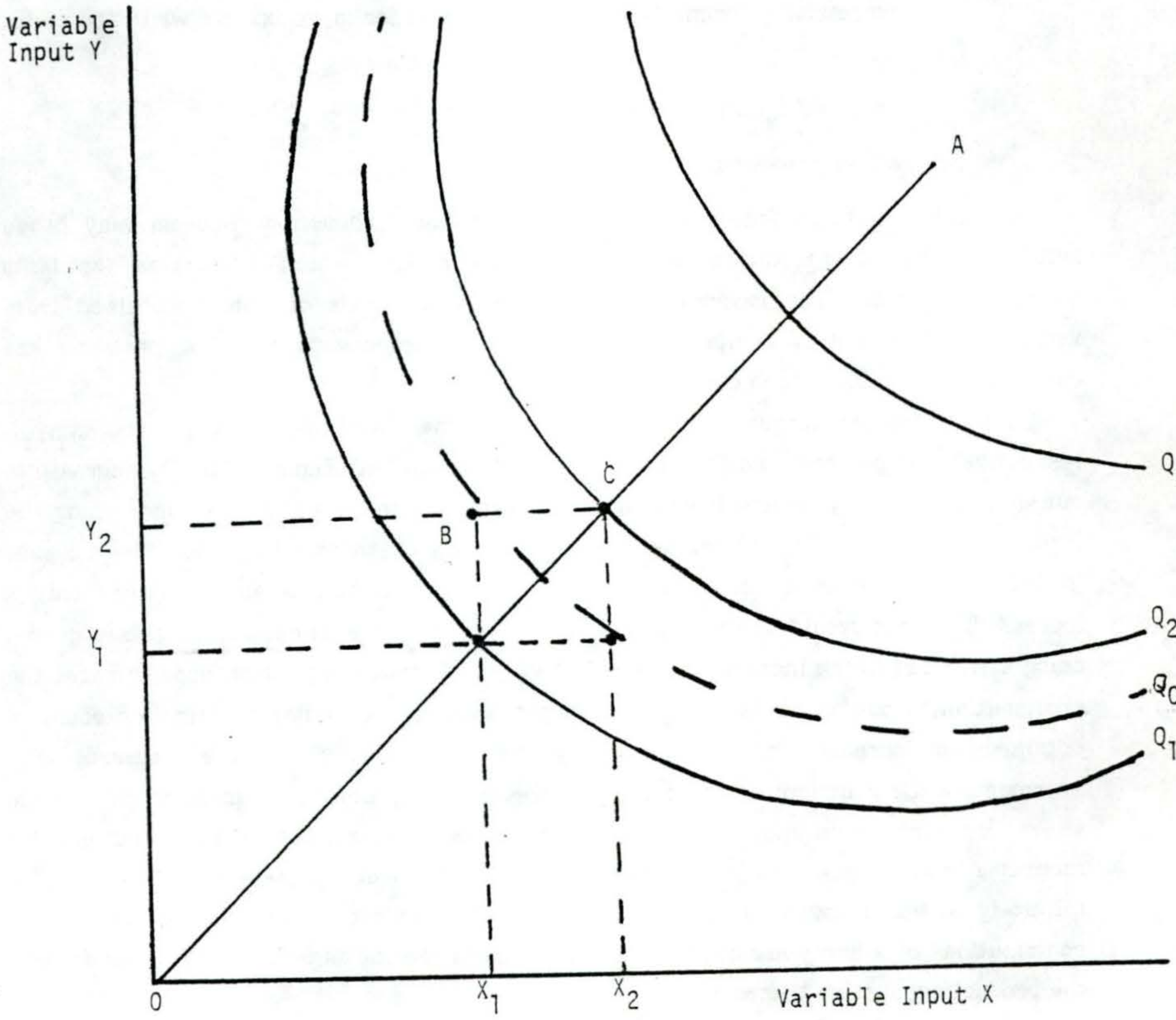


Figure 4.7

A Graphical Presentation of the Product Exhaustion Theorem

At least two observations are worth noting about the relationship between inputs along what is called an expansion path. First, the relationship is independent of the output price. This is because no distinction is made when valuing output between that attributable to x versus that attributable to y. Secondly, any costs associated with the passage of time do not enter into the determination between x and y--it's only those costs that change with use which are affected. For variable inputs that are used up in a single time period, there is no time-related costs. Therefore, the value of the variable input is independent of time--that is determined within each time period. The rule, of course: produce until the value of the output produced from the last input just equals the marginal cost of the input. Thus, the same rule also applies to purchase and use of durable asset services, except that to count all the costs associated with production from a durable requires the introduction of time.

Euler's Theorem

Euler's theorem obtains a result similar in nature to the product exhaustion theorem. It also partitions output among inputs in a way that just exhausts the product. Its limitation is that it applies only to linear homogeneous functions. We now illustrate the product exhaustion formula, obtain Euler's Theorem as a special case, and then provide other examples of the product exhaustion theorem using a homogeneous function not of degree one and a non-homogeneous function.

To begin, a linear homogeneous function measured over input variables x and y has the property:

$$(4.50) \qquad tf(x,y) = f(tx,ty)$$

Interpreted graphically, this definition implies that, for a t percentage increase in both x and y measured along any given linear expansion path, output q will increase by the same percent. It also means that the partial derivatives of f with respect to x and y (f_x and f_y) measured along the expansion path are constants, thus allowing us to describe the output attributed to x and y in a special way. Since f_x and f_y are constants, we obtain from the product exhaustion theorem the result:

$$(4.51) \qquad f_x \int_x dx + f_y \int_y dy = xf_x + yf_y = f(x,y)$$

That $xf_x + yf_y$ equals f(x,y) is Euler's well-known result. That it can be obtained as a special case of the product exhaustion theorem is not so well-known.

We now illustrate the product exhaustion theorem using the linear homogeneous function:

(4.52) $f(x,y) = x^{\alpha}y^{1-\alpha}$

and the relationship between x and y along a linear expansion path as $y = kx$. The unconstrained partial derivative of f with respect to x can be written as:

(4.53) $\partial f/\partial x = \alpha x^{\alpha-1} y^{1-\alpha}$

and substituting kx for y, where kx is the value of y along the expansion path given x, we write:

(4.54) $f_x = \alpha k^{1-\alpha}$

a constant. Similarly, we could obtain an expression for the constant partial derivative for y along the expansion path as:

(4.55) $f_y = (1-\alpha) / k$

As already pointed out, multiplying the constant partial derivatives by the inputs used or integrating returns the same result:

(4.56) $\alpha k^{1-\alpha} \int_x dx + ((1-\alpha) / k) \int_y dy = xf_x + yf_y$

a result which can be graphically portrayed in Figure 4.8. Graphically, the output attributed to x is equal to the rectangle in Figure 4.8, where the horizontal length equals the input level x and the vertical length equals the constant marginal product of x measured anywhere along the expansion path OA in Figure 4.7.

Euler's Theorem holds in this case because the average product equals the marginal product. As a result, multiplying the marginal product of the last unit of production times the total units of inputs used is equivalent to multiplying the average product of x times the total units of inputs used (the area of the rectangle in Figure 4.8)--which obtains exactly that portion of output attributed to x. But whenever the marginal product is not constant, the output attributed to an input cannot be found by merely multiplying the marginal product of the last unit it produced by the total units of inputs used in production. To illustrate, consider two examples where Euler's Theorem does not hold: a non-linear homogeneous function and a non-homogeneous function. By definition, a homogeneous function of degree h has the property that:

(4.57) $t^h f(x,y) = f(tx,ty)$

As an example, consider the non-linear homogeneous function:

(4.58) $f(x,y) = x^{\alpha}y^{\beta}$

where $\alpha + \beta \neq 1$.

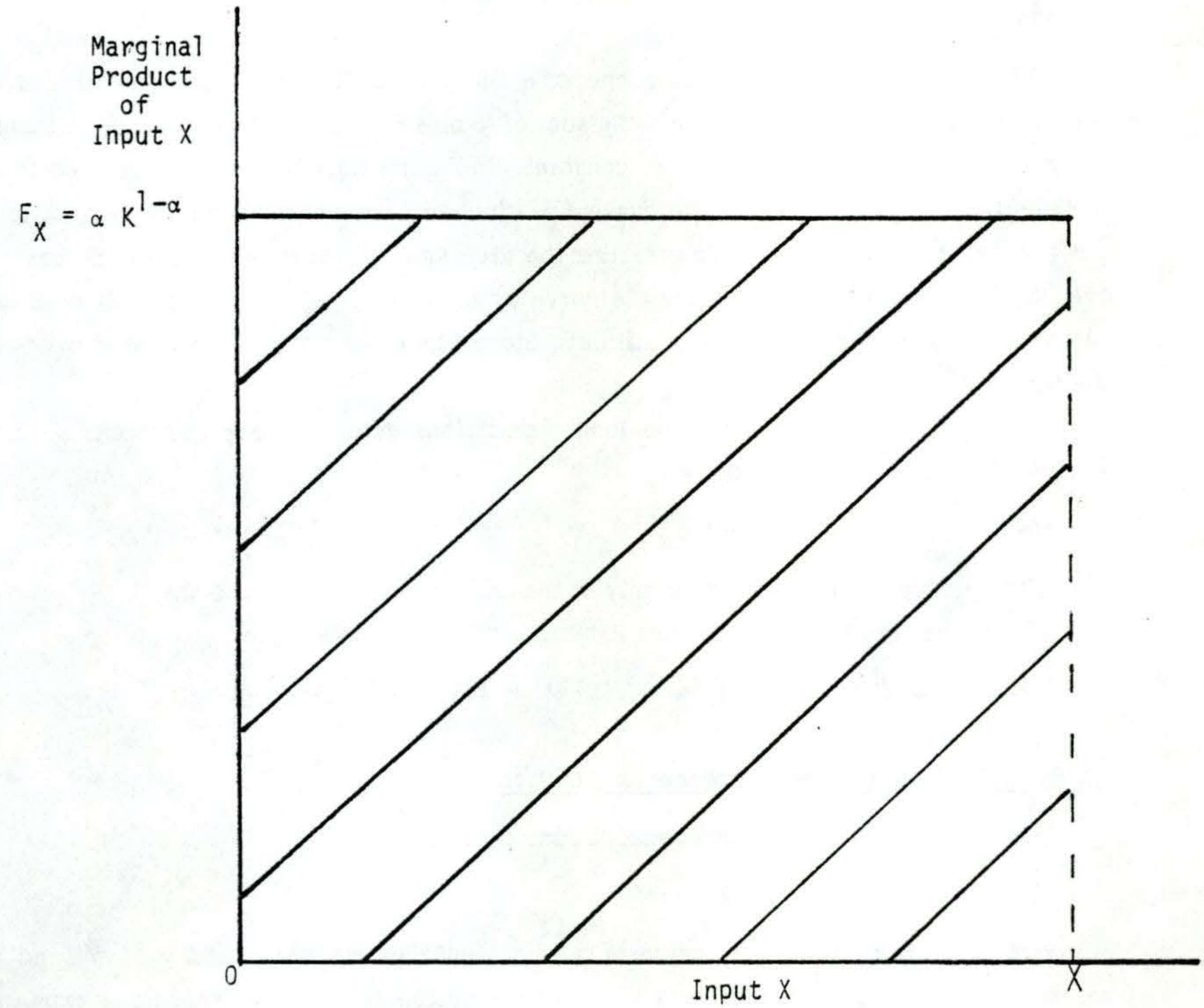


Figure 4.8

Output Attributed to X for a Linear Homogeneous Function
(The Product of X and F_X)

Again, assuming a linear expansion path of the form $y = kx$ allows us to express the derivative of f with respect to x measured along the expansion path as:

$$(4.59) \quad f_x = \alpha k^{\beta} x^{\alpha+\beta-1}$$

Obviously, if α plus β equals one, the function is linearly homogeneous and our earlier results hold. However, when the sum of α plus β does not equal one, f_x measured along the expansion path is no longer constant. In Figure 4.9, the output attributed to x is represented as the area under the curve f_x, which, if $\alpha + \beta > 1$, increases with x (or if $\alpha + \beta < 1$ decreases with x). To measure the area under f_x now requires we integrate f_x over x. To measure the area under the curve by multiplying f_x of the last unit of x used by the units of x used would underestimate the contributions of x if $\alpha+\beta<1$ and overstate their contributions if $\alpha+\beta>1$.

Obviously, Euler's Theorem no longer partitions output among the inputs so as to exhaust the product. To illustrate:

$$(4.60) \quad x f_x + y f_y = \alpha k^{\beta} x^{\alpha+\beta} + \beta y^{\alpha+\beta} k^{-\alpha} = (\alpha+\beta) x^{\alpha} y^{\beta} = (\alpha+\beta) f(x,y)$$

where the product is just exhausted only in the case where $\alpha+\beta$ equals one.

Using the results of the Product Exhaustion Theorem, we write:

$$(4.61) \quad \int_x f_x + \int_x f_y = \alpha K^{\beta} x^{\alpha+\beta}/(\alpha+\beta) + \beta y^{\alpha+\beta} K^{-\alpha}/(\alpha+\beta) = f(x,y)$$

An Example Using a Non-Homogeneous Function

Consider next the non-homogeneous function:

$$(4.62) \quad f(x,y) = x + y - xy$$

Moreover, suppose the output price is p, with input prices for x and y of p_x and p_y, respectively. Then, we could form the profit function from which an output expansion path can be obtained. Letting π be profit, we write:

$$(4.63) \quad \pi = p(x + y - xy) - p_x x - p_y y$$

The expansion path is found by solving for y as a function of x in the expression:

$$(4.64) \quad f_x/f_y = p_x/p_y$$

Taking derivatives of f with respect to x and y in the profit function, we form the above relationship and solve for y to obtain the expansion path:

$$(4.65) \quad y = (p_y - p_x)/p_y + (p_x/p_y)x$$

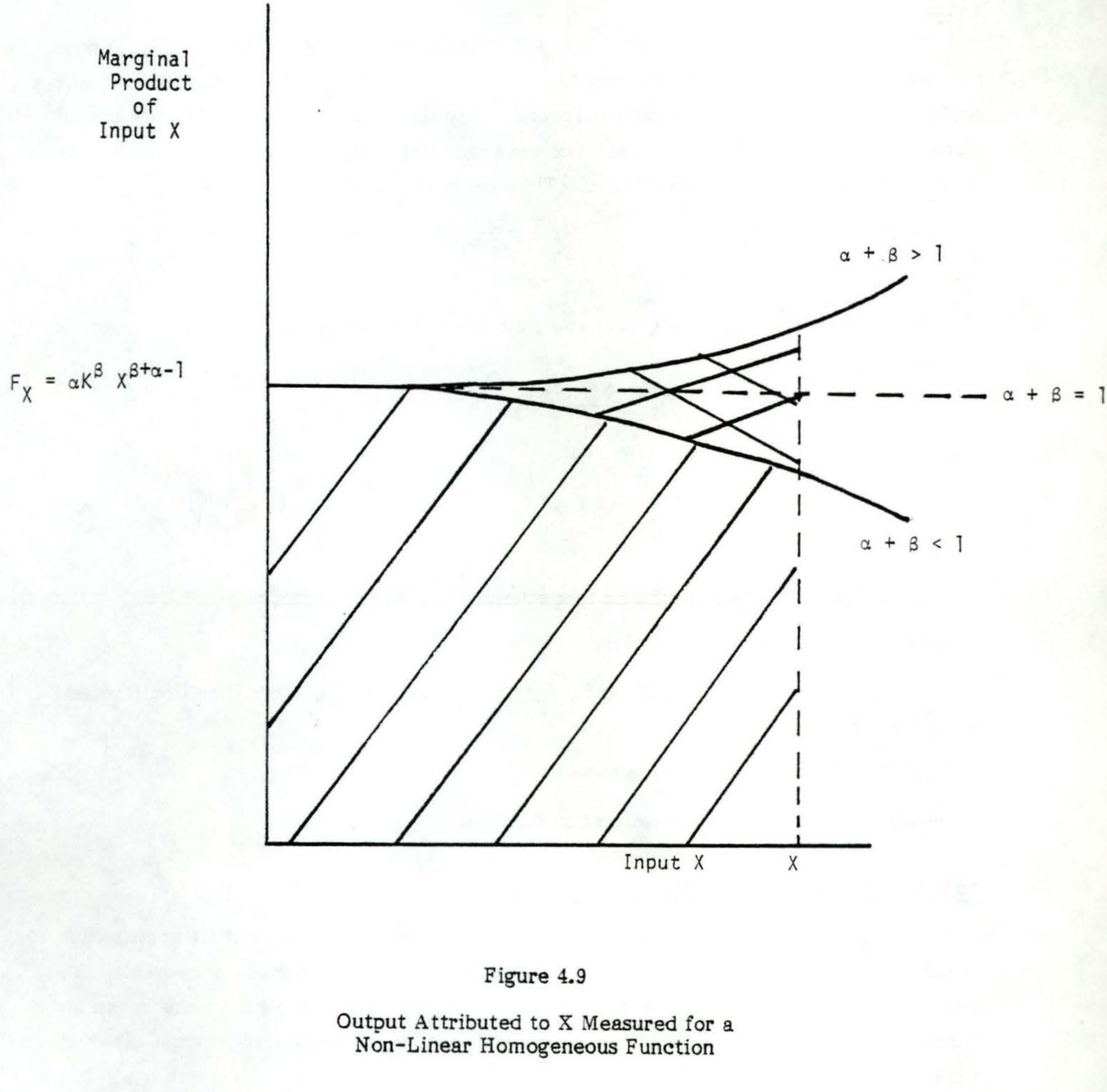


Figure 4.9

Output Attributed to X Measured for a Non-Linear Homogeneous Function

A simpler version would be $y = k_0 + k_1x$ where k_0 and k_1 are constants equal to $(p_y-p_x)/p_x$ and (p_x/p_y), respectively.

The expansion path in this example is illustrated in Figure 4.10. In this case, it does not originate at the origin but begins at the intercept k_0. This complicates the application of the Product Exhaustion Theorem only slightly. Were we to measure output along the expansion path for values of x between 0 and x*, we would not include the output attributed to y between 0 and k_0. Thus, we have two expansion paths:

(4.66) $0 < y < k_0 \qquad x = 0$ and

(4.67) $y = k_0 + k_1x \qquad 0 < x \leq x^*$

The output attributed to y over the first expansion path is k_0; over the second, it equals:

$$(4.68) \quad \int_{k_o}^{y^*} fy\, dy = \int_{k_o}^{y^*} [1- \alpha(\frac{y-k_o}{k_1})]\, dy$$

$$= y^* - (\alpha y^{*2}/2\, k_1) + (\alpha\, k_0 y^*/k_1) - k_0 + (\alpha\, k_0^{\,2}/2\, k_1) - (\alpha\, k_0^{\,2}/k_1)$$

where $y^* = k_o + k_1x^*$.

Similarly, the output attributed to x equals:

$$(4.69) \quad \int_0^{x^*} f_x dx = \int_0^{x^*} (1- y)\, dx = x^* - k_0x^* - (k_1x^{*2}/2)$$

Then, adding together $\int_x f_x dx$ and $\int_x f_y dy$ plus k_o we obtain, after simplification and cancellation:

(4.70) $f(x,y) = x^* + y^* - x^*y^*$

in accordance with the product exhaustion formula.

Applying the Product Exhaustion Theorem (PET)

Having obtained a measurement procedure for ascribing to an input its contribution to output, we are now in a position to answer questions not previously answerable. To best illustrate the power of the new theorem, we compare and contrast the questions answerable with our usual static analysis with those answerable with the Product Exhaustion Theorem results.

Consider a two-input production process f defined over inputs x and y which can be purchased at input prices p_x and p_y. If p is the output price, how much of either x or y should be employed to maximize profit? For this question, our existing static tools are well adapted: acquire until the last unit's cost just equals the value of its marginal product. With diminishing marginal products, we are assured that if the last unit of input

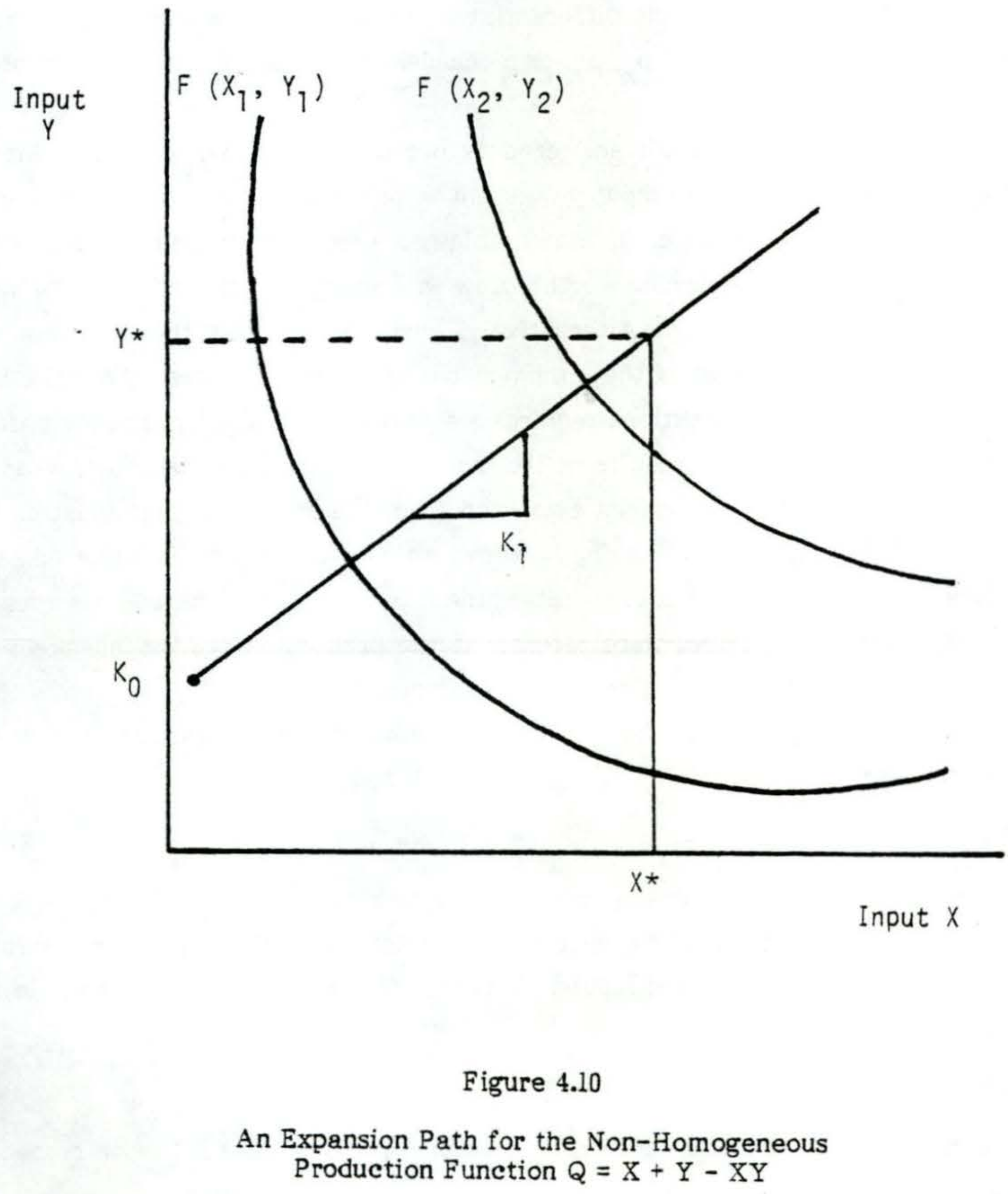


Figure 4.10

An Expansion Path for the Non-Homogeneous Production Function $Q = X + Y - XY$

earns enough revenue to cover its costs, earlier units' returns must exceed their cost. However, whether or not surplus returns exceed non-variable inventory costs is not yet determined.

Once our static economic tools have allowed us to determine the optimal amounts of inputs x and y to employ, implicit differentiation allows us to examine other marginal adjustment questions, such as: if p_x, p_y, or p changes, how will the optimal level of x or y change?

Consider now a question not answered by our usual marginal analysis. Suppose a production process employs (1) an input y that can be purchased and used in divisible units, and (2) services x from a durable. If the durable has a capacity to deliver up to x^* level of services, how much is the durable worth? How will changes in the price of the variable input alter the durable's value? Alternatively, what is the most the firm can pay to acquire the durable? Or, what is the minimum output required to justify purchasing the durable? To answer these questions requires we answer the question ignored before: if the value of the last unit of service from the durable equals its variable cost, what is the value of all previous units of services extracted from the durable? That question is now answerable. But that is not all. Having obtained an expression for the value of services from a durable, we can differentiate to determine how the value of durable services would change if input or output prices were altered; if the production process changed; if the price of the output changed; etc. Consider an example.

Let the production function for output q be the one described earlier, the non-linear homogeneous function:

$$f(x,y) = x^{\alpha}y^{\beta} \tag{4.71}$$

where x and y are inputs and variable costs for inputs x and y are $p_x x$ and $p_y y$, while total returns equal pq. In addition, let FC represent inventory and other fixed costs associated with the durable. Having defined inputs, outputs, and prices, we form the single period profit function π, equal to:

$$\pi = px^{\alpha} y^{\beta} - p_x x - p_y y - FC \tag{4.72}$$

Finding first the expansion path relationship, we differentiate π with respect to x and y and solve for y as a function of x, obtaining first the relationship:

$$\pi_x / \pi_y = p_x / p_y \tag{4.73}$$

from which we obtain:

$$y = p_x \beta x / p_y \alpha \quad \text{or} \tag{4.74}$$

$$y = kx \quad \text{where} \quad k = p_x \beta / p_y \alpha$$

Desiring to know the value of services x extracted from the durable requires we first know the amount of services actually used. Our usual static results answer this question. Using the relationship obtained in (4.74) along with the result that $\pi_x = p_x$ allows us to find the optimal x, $\bar{x}$, such that:

$$(4.75) \quad \bar{x} = (p_x / \alpha p k^{\beta})^{\frac{1}{\alpha+\beta-1}}$$

Next, we form an expression which defines that portion of the output attributable to x. This expression, we know from the PET, equals:

$$(4.76) \quad \int_0^{\bar{x}} \pi_x(x,kx)\,dx \qquad \text{if} \quad \bar{x} \leq x^*$$

$$\int_0^{x^*} \pi_x(x,kx)\,dx \qquad \text{if} \quad \bar{x} > x^*$$

In our example, we can write:

$$(4.77a) \quad \int_0^{\bar{x}} \pi_x(x,kx)\,dx = \int_0^{\bar{x}} (\alpha k\ p_x^{\alpha+\beta-1} - p_x x)\,dx$$

$$= \alpha k\ p\bar{x}^{\alpha+\beta} / (\alpha+\beta) - p_x x$$

Having solved for the optimal level of services to extract from the durable, and knowing the expansion path relationship between x and y, we can write the optimal value of y as:

$$(4.77b) \quad \bar{y} = k\bar{x}$$

Interpreting this result, we have asked the question: what are the net returns attributable to durable services? To answer that question, we first answered the usual static economic question of: what level of service extraction will maximize returns attributable to the durable? The answer was: extract services until the variable cost of extracting the last unit of service equals the marginal value product of the last unit of service extracted. In our example, the expression for $\bar{x}$ in fact represented the derived demand for $\bar{x}$.

Since the production plant, the durable asset which produces services x, is acquired in a lumpy amount, however, we are interested in the total value of services extracted from the durable. Moreover, we wish to know how the value of durable services is affected by changes in variable input prices, output price, and changes in production technology.

We derive such analytic results by letting the expansion path relationship between x and y be written as:

$$(4.78) \quad y = m(x, p_x, p_y)$$

which recognizes explicitly the role of input prices on the expansion path. Moreover, let the derived demand for x be written as:

$$(4.79) \qquad \bar{x} = h(p_x, p_y, p) \quad \text{where}$$

$$d\bar{x}/dp_x < 0 \quad \text{and} \quad d\bar{x}/dp_y \gtrless 0$$

depending on whether x and y are substitutes or complements. Then, the value of durable services V(x) can be written as:

$$(4.80) \qquad V(\bar{x}) = \int_0^{\bar{x}} pf_x[x, m(x, p_x, p_y)]\ dx$$

To find the impact of an increase in p_x on $V(\bar{x})$, we differentiate the above expression to obtain:*

$$(4.81) \qquad dV(\bar{x})/dp_x = pf_x[\bar{x}, m(\bar{x}, p_x, p_y)]\ \partial\bar{x}/\partial p_x + \int_0^{\bar{x}} pf_{xy}[x, m(x, p_x, p_y)]\ (\partial m/\partial p_x)\ dx$$

The first expression on the right-hand side of the above equation gives the impact on the change in the value of services attributed to the durable of a change in the level of services extracted. In this case, the change is negative since $\partial\bar{x}/\partial p_x$ is negative. The second expression gives the change in the value of services resulting from following a different expansion path. In this case, with two variables, f_{xy} of necessity is positive, and the slope of the isoquant increases with increases in the variable price of durable services. Thus, the sign of $dV(\bar{x})/dp_x$ is indeterminate.

Another analytic result could be obtained for changes in $V(\bar{x})$ associated with output price changes, i.e., changes in p. The result is:

$$(4.82) \qquad dV(\bar{x})/dp = pf_x[\bar{x}, m(\bar{x}, p_x, p_y)]\ \partial\bar{x}/\partial p + \int_0^{\bar{x}} pf_x[x, m(x, p_x, p_y)]\ dx > 0$$

since $\bar{x}/\ p$ exceeds zero.

Consider another variant of the problem. Suppose the firm is committed to an output level $\bar{q}$ and experiences an increase in variable input price for the durable services. That is the circumstances facing utilities--they are required to meet demand but may experience variable price increases without a corresponding increase in output price, at least in the short run. To explore the analytic results of such a circumstance, let the

---

*The formula used to find the first-order conditions of an integral can be written as:

$$\frac{d}{dx}\int_p^q f(s,x)\ ds = \int_p^s \frac{\partial}{\partial x}[f(s,x)ds] + f(q,x)\frac{dq}{dx} - f(p,x)\frac{dp}{dx}$$

isoquant relationship, that is, combinations of x and y that produce the same output $\bar{q}$, equal:

$$(4.83) \quad \bar{q} = f(x,y) \quad \text{or} \quad y = f^{-1}(\bar{q},x)$$

To find the optimal amount of durable services to extract, we set $f^{-1}(\bar{q},x)$ equal to $m(x,p_x,p_y)$ and obtain:

$$(4.84) \quad \bar{x} = g(p_x,p_y,\bar{q})$$

Then, the value of services attributable to x equals:

$$(4.85) \quad V(\bar{x}) = \int_0^{\bar{x}} p\, f_x[x,m(x,\, p_x,\, p_y)]\; dx$$

That is $\bar{x}$ is the level of services at the point of intersection between the expansion path $y = m(x,\ p_x,\ p_y)$ and isoquant $\bar{q}$ (see Figure 4.11). Further, as the slope of the expansion falls with increases in p_y or rises with increases in p_x, we obtain:

$$(4.86) \quad d\,\bar{x}/dp_y > 0 \quad \text{and}$$

$$d\,\bar{x}/dp_x < 0$$

The impact on $V(\bar{x})$ of changing p_x or p_y can then be written as:

$$(4.87) \quad dV(\bar{x})/dp_x = pf_x[\bar{x},m(\bar{x},p_x,p_y)]\ \partial\bar{x}/\ \partial p$$

$$+ \int_0^{\bar{x}} \{pf_{xy}\,[x,m(x,p_x,p_y)]\ \partial m/\ \partial p_x\}\ dx$$

Again $\partial\,\bar{x}/\partial p_x$ is negative while $\partial\, m/\partial p_x$ is positive, leaving the sign of (4.87) indeterminate for the reasons already described.

Still other questions not usually asked can be answered using the results of the PET. For example, we may wish to know the level of nondurable input y that leaves returns from the durable just equal to its cost. The answer to this question can be found by solving the expression:

$$(4.88) \quad \int_0^{m^{-1}(y,p_x,p_y)} \{pf_x\,[x,m\,(x,p_x,p_y)]\ - c(x)\,\}\,dx - FC = 0$$

where c(x) is the variable cost associated with extracting services from the durable.

The earlier introduced PET theorem was proven for the case of two inputs. However, by mathematical induction, if we can show that if it holds for the case of three inputs, then it holds for the case of n inputs. Therefore, we introduce the new theorem for the case of three inputs.

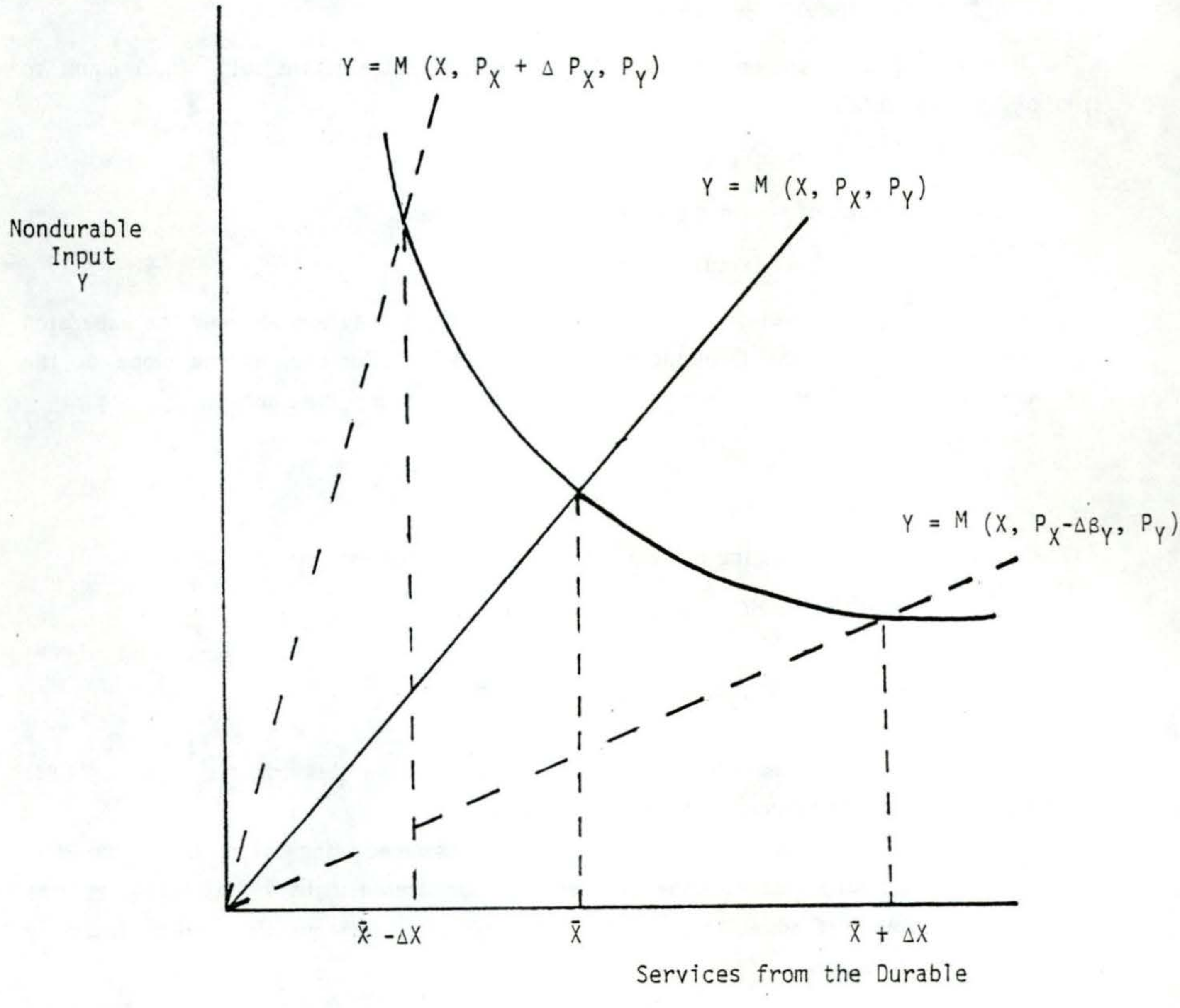


Figure 4.11

A Graphic Description of Optimal Level of Services X to Extract when Output is Fixed and Input Prices Change

The Product Exhaustion Theorem for Three Inputs

Theorem 2. Suppose a general production process is a continuous process defined over arguments x, y and z. That is, let $w = f(x,y,z)$. Then, if the following binary relationships exist: $x = m(y)$, $y = g(z)$ and $z = h(x)$,

$$f(x,y,z) = \int_x f_x\,(x,m^{-1}(x),h(x))dx + \int_y f_y\,(m(y),y,g^{-1}(y))\,dy$$

$$+ \int_z f_z\,(h^{-1}(z),g(z),z)\,dz$$

Proof: Substitute into $f(x,y,z)$, $m^{-1}(x)$ for y and $h(x)$ for z. Then, the derivative of f with respect to x after the substitution equals:

$$f_x = f_x dx + f_y m_x^{-1}(x)dx + f_z h_x(x)\,dx$$

By the fundamental law of calculus, the sum of the anti-derivatives of the partials equals the original function f. That is:

$$\int_x f_x dx + \int_x f_y m_x^{-1}(x)dx + \int_x f_z h_x(x)dx = f$$

By the change-of-variable technique, the following equalities hold:

$$\int_x f_y\,(x,m^{-1}(x),h(x))m_x^{-1}(x)dx = \int_y f_y\,(m(y),y,g^{-1}(y))\,dy$$

and

$$\int_x f_z\,(x,m^{-1}(x),h(x))\,h_x(x)dx = \int_z f_z\,(h^{-1}(z),g(z),z)\,dz$$

Then, after making the required substitutions, we can write:

$$\int_x f_x dx + \int_y f_y dy + \int_z f_z dz = f$$

Completing the Example

The theoretical foundations have been laid to answer all the investment/disinvestment-related decisions for the firm which extracts services from a single durable. The within-period decision depends on variable costs--costs which in turn depend on the relationship between the durable's operating capacity and lifetime capacity, and the market price of the remaining lifetime capacity. Then, benefitting from a new theorem, services from an input can be valued by integrating the constrained partial derivative over the amount of the input employed. Once the value of services is measured, net returns which account for time costs determine the economic life of the durable, with the disinvestment to be made when the returns in the last period equal the annualized average of the net returns in previous periods. This rule is, of course, subject to the provision that lifetime capacity never falls below zero. In practice, the equality either is not always met, or it may be met in every period, or not at all.

Throughout this paper, the investment/disinvestment problem has been illustrated by a simple example. To complete the analysis under certainty and to illustrate the stepwise solution procedures this paper has outlined, we complete our example--see equation (4.24).

STEP 1

Let s, the period for which the durable is held, be two and let output be generated from durable services u(t) combined with a variable input x(t). Then, form the multiperiod gain function G equal to:

$$(4.89)\quad G = pu(1)\,(a/b)^{1/2}\,(1+r)^{-1} + pu(2)\,(a/b)^{1/2}\,(1+r)^{-2}$$
$$-\,[\,(1+u^2(1))\,a + (r+d+rd)\,t^*2a\,]\;\bar{P}\,(1+r)^{-1}\,(1+d)^{-1}$$
$$-\,[\,(1+u^2(2))\,a + (r+d+rd)\,F(2)\,]\;\bar{P}\,(1+r)^{-2}\,(1+d)^{-2}$$
$$-\,p_x\,u(1)^2\,(a/b)^{1/2}\,c\,(1+r)^{-1} - p_x\,u(2)^2\,(a/b)^{1/2}\,c\,(1+r)^{-2}$$

$$\text{such that } F(t) = t^*\,2a - \sum_{j=1}^{t-1}\,[1+u^2(j)]$$
$$u(j) \geq 0 \qquad \forall j = 1, \ldots, s$$

where $\bar{P}$ is the average price of a unit of service--see equation (4.3).

STEP 2

Differentiating with respect to choice variables u(1) and u(2), we obtain deterministic solutions for optimal u*(1) and u*(2) equal to:

$$(4.90)\quad u^*(1) = p(a/b)^{1/2}\,/\,2\,\bar{P}\,[a(1+d)^{-1} - (r+d+rd)\,(1+d)^{-2}\,(1+r)^{-1} + p_x c(a/b)^{1/2}\,\bar{P}^{-1}]$$

$$(4.91)\quad u^*(2) = p(a/b)^{1/2}\,/\,2\,[\bar{P}(1+d)^{-2} + p_x c\,(a/b)^{1/2}]$$

and, since x(t) and u(t) have a deterministic relationship, we can write:

$$(4.92)\quad x^*(1) = u^{*2}(1)\,c(a/b)^{1/2}$$

$$(4.93)\quad x^*(2) = u^{*2}(2)\,c(a/b)$$

STEP 3

Were more than just durable services used in producing an output, Step 3 would identify returns to the durable in each time period by integrating the constrained partial derivative of G with respect to u(t) over u*(t) in accordance with the PET. The result would be a net return for each period. Since, in our model, only the durable is used in the production process, we write the net returns in periods 1 and 2 as:

$$(4.94) \quad g(1) = pu^*(1)\,(a/b)^{1/2} - [\,1 + u^{*2}(1)a + (r+d+rd)\,t^*2a\,]\ \bar{P}\,(1+d)^{-1} - p_x\,u^{*2}(1)\,c(a/b)$$

and

$$(4.95) \quad g(2) = pu^*(2)\,(a/b)^{1/2} - [\,1 + u^{*2}(2)a + (r+d+rd)\,F(2)\,]\ \bar{P}\,(1+d)^{-2} - p_x\,u^{*2}(2)\,c(a/b)$$

STEP 4

Having now determined g(1) and g(2), we are prepared to answer the question: is it time to disinvest? Compare the expression below:

$$(4.96) \quad g(2) \geq r\,[g(1)\,(1+r)^{-1} + g(2)\,(1+r)^{-2}]\ /\ [1 - (1+r)^{-2}]$$

If the expression is true, g(2) exceeds the annualized average of g(1) plus g(2) and should not be replaced. If false, the process returns to Step 1 and considers an alternative value for s.

STEP 5

After determining optimal utilization and life of the durable, and the analysis has been completed for all relevant durables, then select the one with the largest annualized return over its economic life. If positive, invest.

Special Topics

Having established a generalized procedure for solving durable investment/disinvestment and use decisions, we now focus on two issues of interest to utilities which have implication for our model results. The first one is whether or not the rate base--the investment in durables--should be based on replacement or book value. The results may be surprising. The issue is: how the model can be adopted to conditions of required service levels. That is, suppose an output q is required. In this case, the utility company does not have an option of deciding what level of service to provide. To complicate matters further, this required level of demand likely varies.

In this situation, the utility's control variables are the size and type of durable, and when to invest and disinvest. An approach to this problem will be discussed.

Inflationary Impacts on Control Costs

The analysis thus far has ignored inflationary impacts on the model and the resulting optimum decisions. Output prices and input prices were assumed constant; and, while the average price of the durable declined, it was assumed to be as a result of new technology and use. Moreover, the control rate and time discount rate r were assumed to be time preference rates, not subject to inflation.

Now, introduce an inflation rate i into the model and observe the results. Let the output price p, the input price p_x, the durable price P(t), and the discount rate r all increase by an inflation rate i. Were inflation introduced in this manner, there would be no impact on the decisions made by the firm, because increasing the time discount rate by (1+i) in each period exactly cancels the inflationary impact on output price, input price, and the durable's price P(t).

The control rate is not increased by inflation because capital gains exactly offset the inflationary increase in the control cost. Nevertheless, consideration of the appropriate rate to charge for control cost allows us to consider an issue of some importance to the utilities.

The issue is: should the control cost be charged on the replacement cost of the durable or on its depreciated purchase price? That is, should utilities be allowed to earn a return on the replacement or book value of their assets? Our results show three equivalent methods, any one of which should be permitted.

Consider the first of three methods for charging control costs under inflation. For illustrative purposes, we exemplify the results by assuming the firm has acquired a durable of value V which earns a constant real return R, both of which inflate by rate i. Moreover, let the durable's life be n periods and the time preference rate be r. Then, method (1) charges a control cost rate of r on the replacement (inflated) cost of the durable but ignores capital gains that accrue to the firm. The present value of cost C, then, equals:

$$(4.97) \qquad C(1) = r\,V\,(1+i) \,/\, (1+i)\,(1+r) + \ldots + r\,V\,(1+i)^n \,/\, (1+i)^n\,(1+r)^n$$

$$= V\,[\,1 - (1+r)^{-n}\,]$$

The second method charges an inflated control rate of r+i+ri times the book or purchase price of the durable. The result is:

$$(4.98) \qquad C(2) = (r+i+ir)\,V \,/\, (1+i)\,(1+r) + \ldots + (r+i+ir)\,V \,/\, (1+i)^n\,(1+r)^n$$

$$= V\,[\,1 - (1+r)^{-n}\,(1+i)^{-n}\,]$$

The third method charges an inflated control cost rate r+i+ir on the inflated durable, but reduces cost by the amount of capital gains.

$$(4.99) \quad C(3) = (r+i+ir)\ V\ (1+i)\ /\ (1+i)\ (1+r) + \ldots + (r+i+ir)\ V\ (1+i)^{n}\ /\ (1+i)^{-n}\ (1+r)^{-n}$$
$$- (i+ir)\ V\ (1+i)\ /\ (1+i)\ (1+r) - \ldots - (1+ir)\ (1+i)^{n}\ V\ /\ (1+i)^{n}\ (1+r)^{n}$$
$$= V\ [\ 1 - (1+r)^{-n}\ ]$$

The present value of costs C(1) exactly equals the present value of costs C(3) net of capital gains. Further, while C(2) differs slightly from C(1) and C(3), in the limit, when n is large, the difference is not significant. So, we conclude that allowing utilities to charge control costs based on an inflated control rate applied to book value is an acceptable alternative to being allowed to charge a control rate discounted for inflation applied to replacement cost of the durable.

Meeting Time-Varying Required Levels of Output

The second special topic considers how our analysis to date would be changed by the firm not being allowed to choose a desired level of output but rather being required to meet the demand. Let that time-varying level be:

$$(4.100) \quad Q(t) = u(t)\ (a/b)^{1/2}$$

The multi-period gain function could then be written as:

$$(4.101) \quad G = pQ(1)\ (1+r)^{-1} + \ldots + pQ(s)\ (1+r)^{-s}$$
$$- \{\ [a+bQ^{2}(1) + (r+d+rd)\ t^{*}2a]\ \ (1+d)^{-1}\ \bar{P} + p_{x}(b/a)^{1/2}\ Q^{2}(1)\ c\ \}\ (1+r)^{-1}$$
$$- \ldots$$
$$- \{\ [a+bQ^{2}(s) + (r+d+rd)\ F(s)]\ \ (1+d)^{-2}\ \bar{P} + p_{x}(b/a)^{1/2}\ Q^{2}(s)\ c\ \}\ (1+r)^{-s}$$

such that:

$$F(t) = t^{*}2a - \sum_{j=1}^{t-1} [a+bQ^{2}(t)]$$

and $F(t) \geq 0$ for t = 1, . . ., s

Then, if in fact "a" and "b" are choice parameters determining the rated capacity and life of the durable, the multi-period gain function restricted to meet demand levels Q(t) can be differentiated and solved. Since the derivatives of G with respect to "a" and "b" are not simple and lend to no intuitive information, except that they depend on weighted values of Q(t), they are not reported here. Obviously, analysis involving more periods would add additional weights to consider, weights which depend, of course, on the required levels of outputs Q(1), Q(2), . . . These problems, because they become quite complicated, can best be solved with the aid of a computer.

Including the Cost of Uncertainty

Into the previous analysis, we now include another consideration--uncertainty. Uncertainty produces different responses by decision makers, depending on their marginal utility for income. Assume that, by applying the expected utility hypothesis, we have identified a decision maker's utility function U, a function that exhibits diminishing marginal utility so that $U' \geq 0$ and $U'' < 0$ (see Figure 4.12).

The cost of risk to a decision maker can be defined and measured using the utility function U. Assume the decision maker, whose utility function is illustrated in Figure 4.12, owns an investment which will, with equal probability, earn $w - x$ or $w + x$. We propose, as others do, to define the cost of risk for this decision maker as the amount the decision maker would subtract from his expected income to eliminate uncertainty. Note that the utility of expected income plus wealth w, for concave utility function U(w), is greater than the expected utility of the gamble represented as EU(w), so that our decision maker would pay some positive amount to eliminate the uncertainty. In particular, he would pay an amount π since, for that amount, he is indifferent between the certain income $w - \pi$, valued at $U(w - \pi)$, and the expected utility of income plus wealth, equal to $1/2\ U(w - x) + 1/2\ U(w + x)$. The amount π we call the cost of risk.

Note the important relationship between the cost of risk π and the curvature of U, measured by the negative ratio of $-U''/U'$. As long as U" is negative, the function is concave and π is positive. In general, as U" approaches zero, that is, U approaches the straight line AB, the cost of risk diminishes. So, for decision makers who have constant marginal utility for income, which is what a linear utility function infers, there is no cost of risk. In fact, should $U'' > 0$, the decision maker would pay a premium to assume the gamble. The important lesson is that the cost of risk depends on decision makers' attitudes towards additional income, as well as on the uncertainty of the action choices.

To measure the cost of risk under more general conditions than the two-outcome uncertainty case described earlier, let x represent uncertain outcomes of a gamble with expected value 0 and variance σ^2. Then, if we have measured the cost of risk accurately, the decision maker will be indifferent between the expected utility of the action choice with uncertain income and the action choice that earns $w - \pi$ with certainty. We express this indifference as:

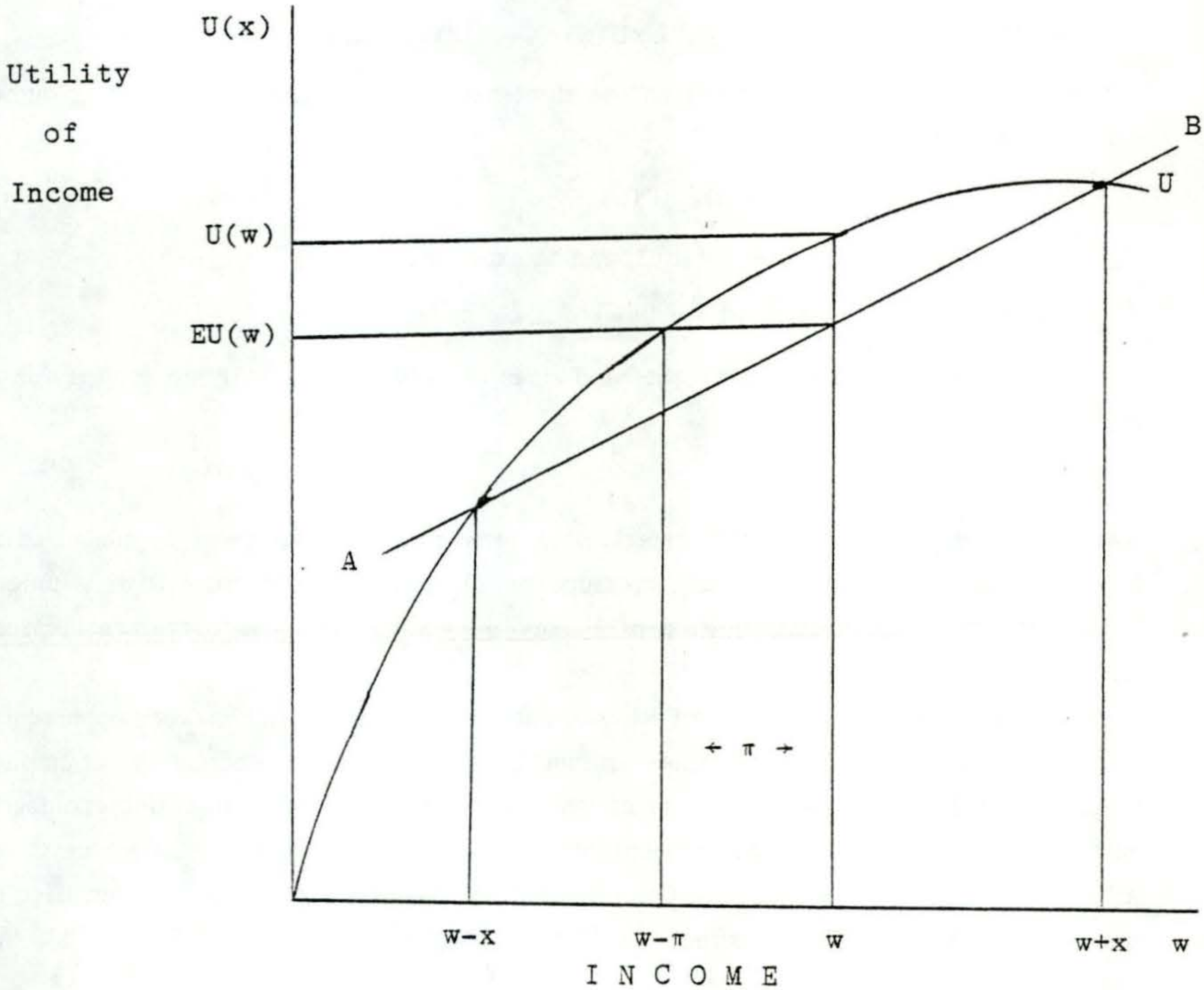


Figure 4.12

An Illustration of How the Curvature of a Utility Function Affects the Cost of Risk

(4.102) $EU(w + x) = U(w - \pi)$

To obtain a closed form expression for π, we follow Pratt (1964) and expand both sides of (4.102) around π and x using a Taylor polynomial and ignoring higher-order terms.

The left-hand side equals:

(4.103) $$EU(w + x) = E[U(w) + xU'(w) + \frac{x^2}{2} U''(w) + ...]$$

and, after finding the expectations on the right-hand side of (4.103) and ignoring higher moments, we can write:

(4.104) $$EU(w + x) = U(w) + U''(w) \frac{\sigma^2}{2}$$

Next, the right-hand side of (4.102) can be expanded as:

(4.105) $$U(w - \pi) = U(w) - \pi U'(w) + ...$$

Finally, after equating the right-hand sides of (4.104) and (4.105) and solving for , we obtain:

(4.106) $$\pi = \frac{-U''(w)}{U'(w)} \frac{\sigma^2}{2}$$

We have now shown explicitly the relationship between the curvature of U, measured by $R(w) = -U''/U'$ and the risk premium measure π. Moreover, the measure R(w) is unique since, for any linear transformation of U, say $U^* = a + b\, U(w)$, the curvature measure R(w) is unaltered.

An additional interpretation of R(w) can be made if the decision maker's opportunity set is described by an expected value-variance (EV) set. Such an opportunity set implies that, for the decision maker, the cost of risk is positive since, between action choices of equal means, he chooses the variance-minimizing choice. Consider the EV opportunity set AB in Figure 4.13. The preferred action choice from the set AB is, of course, identified by the tangency between the isoexpected utility function $\bar{U}'\bar{U}$ and AB. In Figure 4.13, this solution is represented by expected wealth, w_1, and variance σ_1^2.

The solution could also be identified by drawing a linear tangent to the EV set so that slope R(w)/2 equalled the slope at equilibrium between the EV set and the isoexpected utility function. In Figure 4.13, this tangent line identifies expected income as a function of some certain income w_o and variance.

(4.107) $$w = w_o + \frac{R(w)}{2}\sigma^2$$

Note that $w_1 - w_o$, that is, the difference between the expected value of an action choice and the certain income that leaves the decision maker indifferent between the uncertain action choice and the certain income, equals what we defined earlier as the cost of risk. Further, rearranging (4.107), we have:

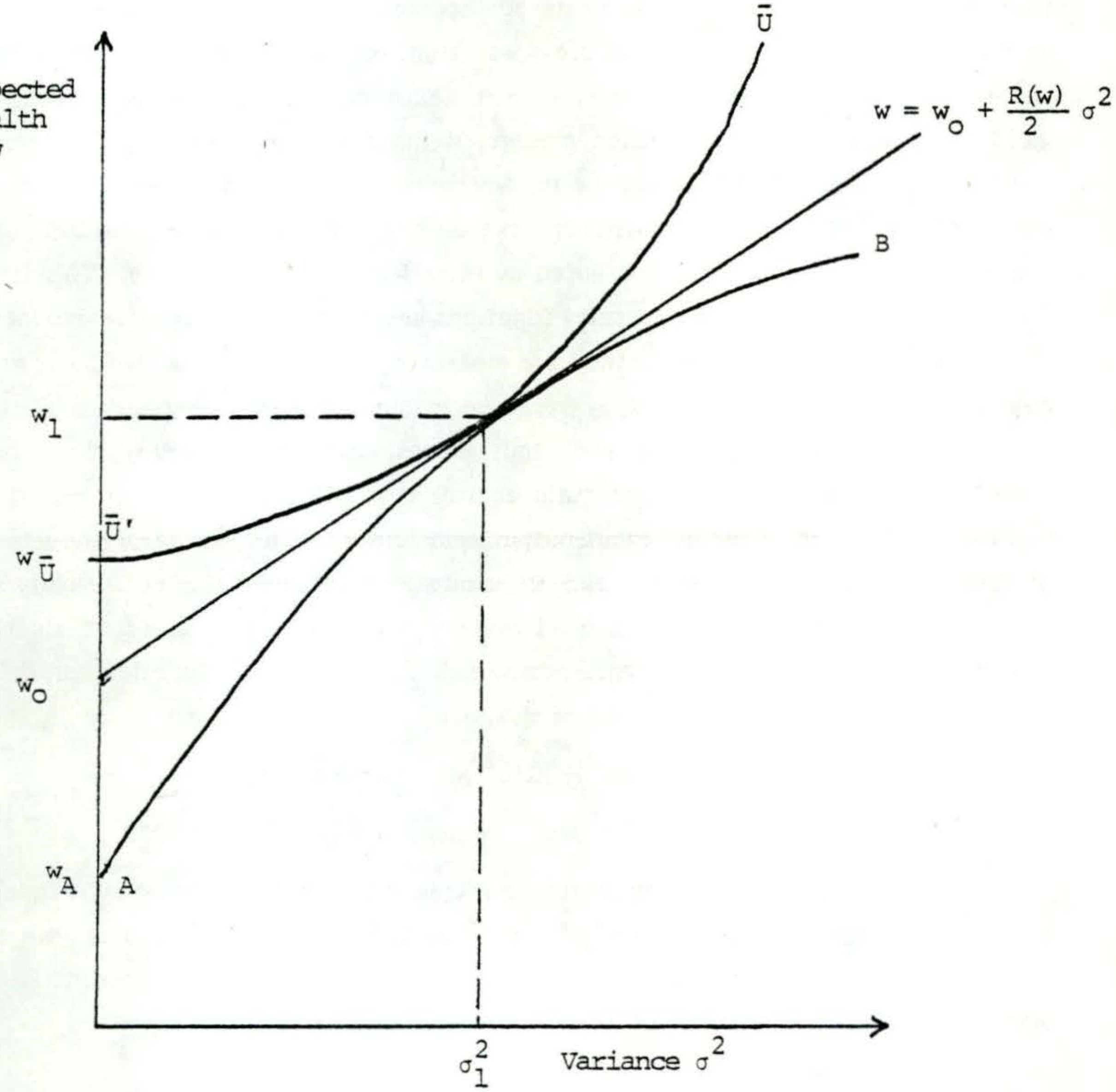


Figure 4.13

Equilibrium between an EV Efficient Set and a Decision Maker's Isoexpected Utility Function

$$(4.108) \qquad w_1 - w_o = R(w)\,\sigma^2/2$$
$$= \pi$$

The expression in (4.108) is, of course, identical to the expression in (4.106), so that our curvature measure $\frac{-U''}{U'}$ can also be interpreted as the equilibrium slope on an EV set. However, it is only an approximate measure. The true measure in Figure 4.13 would be the difference between $w_1 - w_{\overline{U}}$, while our approximate measure is $w_1 - w_o$. Still, our approximation is useful because it provides us an expression for the certainty equivalent value of an action choice. Moreover, we can maximize this certainty equivalent, which is valid for a wide range of decision makers, without knowing the specific form of their utility function--provided we choose the appropriate R(w). As a result, we can modify our objective function G by subtracting the cost of risk. If R(w) is constant, of course, the risk cost approximation represented by (4.108) is an exact measure. This is the case for the class of investors whose utility functions are described by negative exponentials.

Now, having obtained a method for measuring the cost of risk, we ask: what gives rise to uncertainty (represented by variance in our model)? Several possibilities exist. These include uncertain output and input prices, uncertain salvage prices due to an uncertain discount rate d, an uncertain capital cost rate r, an uncertain rated capacity and user cost, or an uncertain input-output relationship. The source of uncertainty will, of course, influence the cost of risk. If output price is uncertain, as is usually assumed, with variance equal to σ^2 and expected value $\overline{p}$, the cost of risk is output squared times the variance of risk, and the single-period risk discounted gain function can be written, with the last term equalling the cost of risk, as:

$$(4.109) \qquad g(j) = \overline{p}u(j)\,(a/b)^{1/2} - [(1+u^2(j))\,a + (r+d+rd)\,F(j)]\,(1+d)^{-j}\,\overline{P}$$
$$- p_x\,u^2(j)\,c\,(a/b)^{1/2} - R(w)\,[u(j)\,(a/b)^{1/2}]^2\,\sigma^2/2$$

Ignoring indirect capacity costs and treating R(x) as a constant--equal to the average risk aversion coefficient of all participants in the decision process--we can solve for the optimal utilization rate in the j-th period, u(j), under risk. Differentiating (4.109) above and solving for u(j), we obtain:

$$(4.110) \qquad u(j) = \overline{p}/\,[2(ba)^{1/2}\,(1+d)^{-j} + 2p_x\,c + R(w)\,(a/b)^{1/2}\,\sigma^2]$$

Note in (4.110) the inverse relationship between positive value of R(w), which measures risk aversion, and the utilization rate u(j). This inverse relationship implies that, if the utilization rate is a decision required before the resolution of uncertainty, then the rate selected will be lower as risk aversion R(w) increases. Similarly, the utilization rate will decrease with increases in rated capacity $(a/b)^{1/2}$ and with increases in the variance of output price σ^2.

We earlier related flexibility to time capacity costs. The larger the loss in capacity due to time, the larger the output at which average losses in lifetime capacity are minimized. This can be demonstrated in our example by letting z(t) be the rated capacity $(a/b)^{1/2}$, and by differentiating with respect to "a," the coefficient determining time capacity losses, to obtain:

$$(4.111) \qquad dz(t)/da = 1/2(ab)^{-1/2} > 0$$

We might ask the question, how will increasing inflexibility, increases in "a," alter the optimal service extraction rate?

On the one hand, increased inflexibility increases the cost of operating at utilization rates other than 100 percent--but the costs are not symmetric. It is more costly to overutilize than to underutilize (because of the quadratic direct user cost function). So, this result alone would suggest the utilization rate would decrease. However, if, on the other hand, a larger portion of the costs are fixed, then the cost of risk will be reduced and the utilization rate will increase. As a result of these opposing forces, it is not surprising that the derivative of u(j) in (4.110) with respect to "a" is ambiguous.

$$(4.112) \qquad du(j)/da = \bar{p}\ [(R(w)\ (a/b)^{1/2}\ a\ \sigma^2/2b^2) - (ba)^{-1/2}\ b(1+d)^{-j}]\ \lesseqgtr\ 0$$

A perhaps more relevant introduction of uncertainty is on the demand side. Instead of uncertain prices, let the demand for services from the durable be a random variable q with expected value $\bar{Q}$ and variance equal to σ_q^2. Then, introducing the constraint that services from the durable, $u(j)\ (a/b)^{1/2}$, must equal demand q, we can write the risk discounted gain function as:

$$\begin{aligned}(4.113) \qquad g(j) &= p\bar{Q} - [\ (a + b\ (\sigma_q^2 + \bar{Q}^2) + (r+d+rd)\ F(j)\ ]\ (1+d)^{-j}\ \bar{P} \\ &- p_x\ (b/a)^{1/2}\ c\ (\sigma_q^2 + \bar{Q}^2) - \frac{R(w)}{2}\ [p^2\ \sigma_q^2 + \{\ b^2\ (1+d)^{-2j}\ p^2 \\ &+ p_x^2\ (b/a)\ c\ \}\ \sigma_q^2 2\ ]\end{aligned}$$

where $\sigma_q^2 2$ is the variance of the squared random variable q.

The cost of risk with random demand is more complicated than our earlier expression for risk costs with random output prices because uncertain demand for durable services affects both costs and income. Income obviously is affected since only the actual demand gets met, even though the firm's capacity to produce services exceeds demand. As a result, capacity costs depend on actual use, which becomes random as demand varies stochastically, and which adds to the cost of risk. Variance of q enters in the expected value portions of the cost of risk expression, because of the squared q term in measuring user cost, and, for the same reason, the variance of q^2 enters in the risk portion of the equation.

If the firm faces random demand, which in the case of utilities must be met, then utilization rates can only be altered by changing the capacity of the durable. If it could be achieved at no extra cost, perfect flexibility would be desired--a constant average loss in lifetime capacity regardless of the demand level. Unfortunately, however, flexibility occurs at a cost. In fact, it becomes somewhat of a definitional problem to determine when one durable is more flexible than another except by reference to the completely flexible durable.

One way to order durables according to their flexibility is to weight their respective average loss curves, e.g., $AL_1(Q)$ and $AL_2(Q)$, by the probability that Q occurs, namely f(Q). Then, in a manner similar to first-degree stochastic dominance (Hadar and Russell, 1969), define $AL_1(Q)$ to be more flexible than $AL_2(Q)$ if:

$$(4.114) \qquad \int_{-\infty}^{Q^*} [AL_1(Q) - AL_2(Q)] \; f(Q)dx \leq 0 \qquad \text{for all } Q^* \in [-\infty, \infty]$$

This criterion, however, would leave unranked a large number of durables. Alternatively, we may wish to rank the durables according to their respective expected average losses. However, this would ignore diminishing marginal utility of income. So, introduce a utility function U defined over Q, U(Q). The result would suggest $AL_1(Q)$ is more flexible than $AL_2(Q)$ if:

$$(4.115) \qquad \int_{-\infty}^{\infty} [U(AL_1(Q)) - U(AL_2(Q))] \; f(Q) \geq 0$$

Although the tradeoffs to obtain flexibility cannot be generally defined, an example may help illustrate the division problem. Earlier, we defined a completely flexible durable as one whose losses in lifetime capacity equalled services extracted. Letting z(t) be services extracted and F(t) - F(t+1) be losses in lifetime capacity, the relationship between losses and services extracted could be written as:

$$(4.116) \qquad F(t) - F(t+1) = a + b\, z^c(t)$$

If "a" is zero and "b" and "c" are 1, the durable is completely flexible (see line OA in Figure 4.1).

Increasing time capacity costs (increasing "a" in our example) or increasing marginal losses in lifetime capacity (increasing "b" or "c" in our example) would both reduce flexibility. Now, suppose the engineers have identified trade-offs in design such that:

$$(4.117) \qquad ab = k \qquad \text{and} \qquad c = 2$$

where k is some constant. That is, reductions in time capacity costs, reductions in "a," can only be achieved by increases in the marginal loss coefficient "b." Figure 4.14 illustrates the extremes with curve AB, which has a minimum time capacity cost, versus curve CD, which has a higher time capacity cost but a lower marginal loss coefficient.

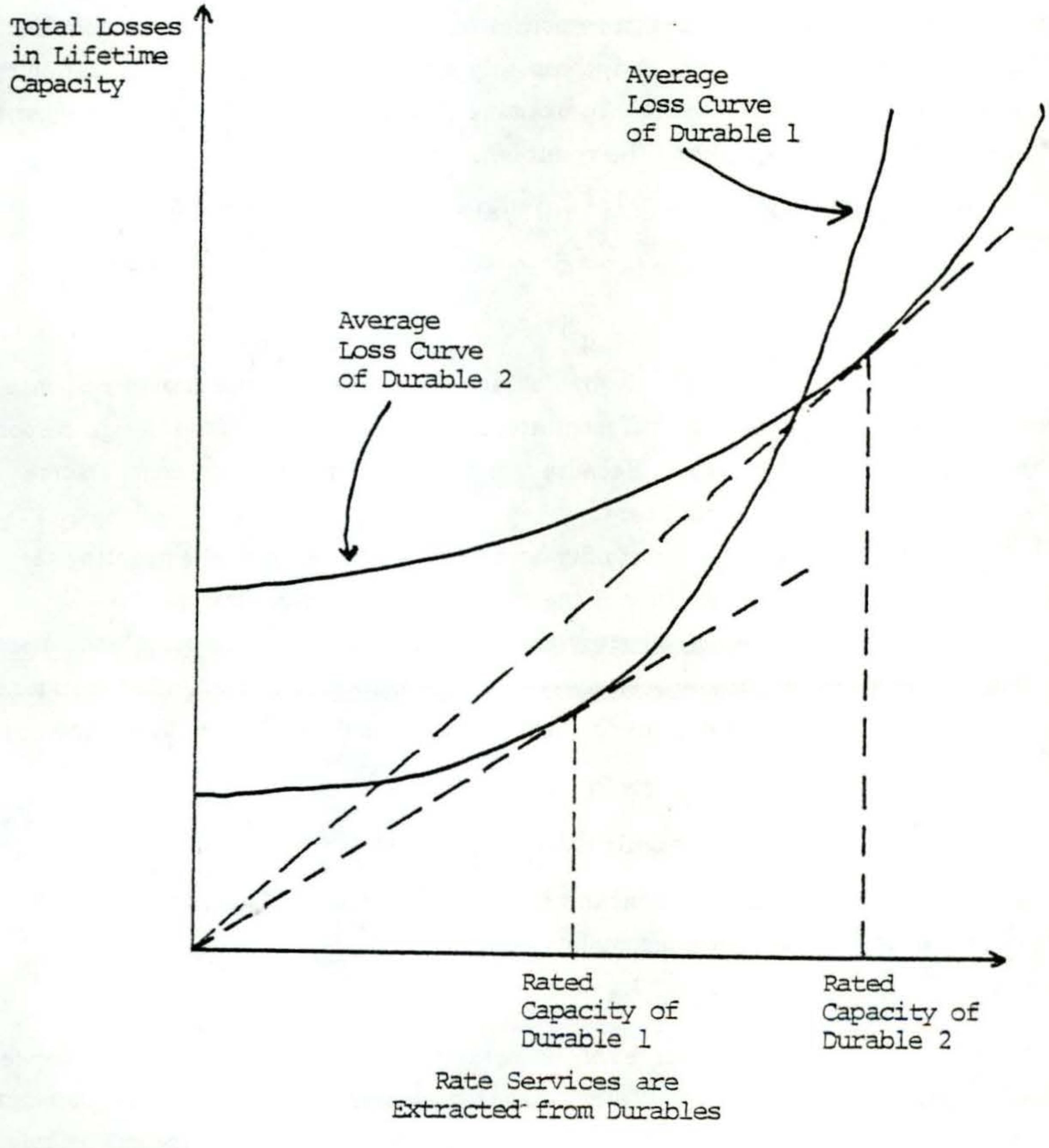


Figure 4.14

Relationships between Time Capacity Costs, Marginal Loss Coefficients, Rated Capacities, and Flexibility

Increasing "a" increases the rated capacity but in the process reduces flexibility for output levels less than the rated capacity.

If our constraint can be considered as an example of tradeoffs in design, then it might be useful to explore the design responses for the firm which faces random demand for its output which must be met and whose only alternative to adjusting utilization rates is to alter the design of the durable. To examine this feature of the model, we substitute for b in (4.113) the expression k/a. The result is:

$$(4.118) \quad g(j) = p\bar{Q} - [(a + (k(\sigma_q^2 + \bar{Q}^2)/a) + (r+d+rd)F(j)] (1+d)^{-j} \bar{P} - p_x k^{1/2} c(\sigma_q^2 + \bar{Q}^2)/a - \frac{R(w)}{2} [p^2 \sigma_q^2 + k^2 (1+d)^{-2j} \bar{P}^2 + p_x^2 (k/a^2) c \sigma_q^2 2]$$

Then, differentiating implicitly by "a" and R(w), we find the design response for increased risk aversion. Similar differentiations would produce optimal design responses to increased uncertainty, etc. Because they are ambiguous and cumbersome, the differentiations are not presented here.

Finally, we consider the uncertainty associated with the cost of extracting services from the durable and the uncertainty of the remaining lifetime capacity.

The relationship between direct user cost and time capacity costs and losses in lifetime capacity has up to now been treated as a known relationship. In practice, this relationship is uncertain. An example of this uncertain relationship could be expressed as:

$$(4.119) \quad F(t) - F(t+1) = a + (b + \varepsilon) z^2(t) = a[1+u^2(t)] + (a/b) u^2(t) \varepsilon$$

where is a random variable with mean zero and variance σ_ε^2. Then, it follows that the cost of risk associated with this uncertainty equals:

$$(4.120) \quad \pi = R(x) [u^4(t) (a/b)^2 \sigma_\varepsilon^2]$$

Consider still another source of uncertainty--the remaining lifetime capacity. Suppose that, instead of knowing lifetime capacity, we know the probability distribution of lifetime capacity and that that distribution can be described by the random variable:

$$(4.121) \quad F(t) = F(t) \varepsilon_F$$

where ε_F has expected value of one and variance of σ_F^2. Then, the variance of lifetime capacity can be written as:

$$(4.122) \quad \text{Var}[F(t)] = F^2(t) \sigma_F^2$$

The relationship identified above suggests that, as the expected value of lifetime capacity declines, so does the variance.

Consider an example that typifies the expression above. Assume you are considering the purchase of a used car--one with 10,000 miles on the odometer. The salesman assures you these miles were extracted at the car's rated capacity--still you wonder. If the car is a "lemon," it may only last a few thousand miles; on the other hand, if the salesman is truthful, it may last as many as 100,000 miles. The range of possibilities, which is in direct relationship to the variance, is between 10,000 and 100,000 miles--a very large variance.

Consider the alternative--a 1969 Buick LeSabre with 95,000 miles. How many more miles can be left? Possibly 5,000, not more than 10,000. The range of possibilities is 10,000 miles--a very small range (and variance) of possible remaining lifetime capacities compared to the first car.

The cost of risk associated with uncertain lifetime capacity is primarily associated with time depreciation and control costs (and indirect capacity costs in the multi-period example) because these are directly related to the lifetime capacity measure. As a result, the risk cost in the j-th period associated with uncertain lifetime capacity can be written as:

(4.123) $$\pi = R(x)\,[(r+d+rd)\,F(t)\,(1+d)^{-j}\,\bar{P}]^{2}\,\sigma_F^2$$

These risk costs associated with uncertain remaining lifetime capacity and user cost provide a possible explanation for market prices of durables that decline at decreasing rates rather than at increasing rates as in equation (4.5) for the certainty model: in the process of using up lifetime capacity, the variance of remaining lifetime capacity is simultaneously reduced.

In both risk cost functions, increasing risk aversion and variances reduce utilization rates below their certainty levels, but, because the derivatives are cumbersome and not particularly revealing, we omit them.

From Theory to Practice

So far, we have introduced various ways uncertainty may enter the investment/disinvestment problem. In all cases, uncertainty adds additional costs and alters the optimal service extraction rate. Adding risk costs that are related to use in essence makes more of the costs variable, resulting in the optimal service extraction rates being smaller than is the case where more of the costs are time-related.

Having made this observation, however, we see how quickly our theoretical model can become complicated. As a result, a practical model of any realistic size requires the aid of a computer to solve. Such a computer model is introduced and solved in the subsequent chapters.

GLOSSARY OF TERMS

Capacity Costs:	Those costs occurring because of changes in the capacity of the asset to deliver services.
Capacity Time Cost:	A capacity cost resulting from losses in the durable's lifetime capacity due to the passage of time.
Control Cost:	An inventory cost equal to the rate of return that funds committed to the durable could earn (or cost) in their next best opportunity.
Direct User Cost:	A capacity cost resulting from the loss of the durable's lifetime capacity due to use.
Durable Assets:	Assets which provide services for more than one time period.
Fixed Costs:	Those costs that vary with time but not use.
Indirect Capacity Costs:	Those future-period costs (benefits) resulting from current-period use decisions.
Indirect User Cost:	An indirect capacity cost equal to the savings (costs) resulting from reduced control and time depreciation costs due to use in the current period.
Inventory Costs:	Those costs that occur as a result of holding an inventory of services over time.
Lifetime Capacity:	The total amount of services available from the durable if services are extracted at the durable's rated capacity.
Maintenance Cost:	A capacity cost which modifies the direct user cost or time capacity costs. It can also be considered to be a durable investment if its services extend beyond a single time period.
Non-Durable Assets:	Assets used up in the production process in the current period.
Operating Capacities:	The potential range at which services can be extracted from the durable.
Product Exhaustion Theorem (PET):	A theorem which provides a method for partitioning the total value of a function among the input variables--in such a way that the sum of the outputs attributed to the input variables just equals the total value of the function.

Rated Capacity:	The operating capacity which minimizes the average loss in lifetime capacity.
Replacement Opportunity Cost:	An indirect capacity cost equal to the average returns foregone by postponing replacement one period.
Time Depreciation Cost:	An inventory cost (benefit) which is a result of a re-evaluation of the durable's remaining lifetime capacity by market forces--in addition to those changes in the durable's market value as a result of time.
Variable Costs:	Those costs that vary with use but not time.

CHAPTER V

IMPLEMENTING CHOICE THEORY UNDER UNCERTAINTY

Robert P. King

Introduction

The expected utility hypothesis is the basis for much of the large body of theory concerned with decision making under uncertainty. It states that choices made under uncertainty are affected by a decision maker's preferences for alternative outcomes and by his expectations concerning the relative likelihood of alternative outcomes under each of the action choices being considered, and it is the source of a general decision rule which integrates information on these two factors. Despite its wide acceptance as a theoretical tool, the expected utility hypothesis is often not used in the analysis of practical decision problems in which uncertainty has an important impact. This is due, to a large extent, to several important operational problems which make implementation of decision theory based on the expected utility hypothesis difficult in a practical context. One particularly serious set of problems stems from the fact that implementation of the expected utility hypothesis requires that some determination be made of the decision maker's preferences and expectations. Both have proven to be difficult to measure accurately enough to be a reliable basis for the evaluation of alternative choices. Another set of difficulties can be attributed to the fact that commonly used optimization techniques are not well suited for the analysis of many decisions made under uncertainty. As a result, rather severe restrictions are often imposed on the representation of decision-maker preferences and expectations and on the types of choices considered.

This paper focuses on one of the problems identified above--that of measuring decision maker preferences. Shortcomings of existing approaches to the measurement and representation of preferences are identified, and a new method for measuring preferences is introduced. This procedure was developed as part of an integrated set of techniques which also addresses problems related to the determination and representation of expectations and the identification of preferred choices.

The most direct approach to the measurement of decision-maker preferences is to derive the decision maker's utility function. A utility function is a relationship between the outcome of a choice and an index of its desirability. It assigns values to alternative situations or conditions. As such, a utility function is a highly structured representation of a decision maker's preferences. Unfortunately, due to shortcomings in the design of preference elicitation interviews (Officer and Halter, 1968), problems in statistical estimation (Knowles, 1980), and respondents' own lack of precise knowledge about their preferences (Zadeh, 1973), utility functions also tend to be unreliable representations of preferences. Despite such problems, utility functions, once estimated, are usually treated as though they were exact representations of preferences. When alternative choices are ordered, any absolute difference in the expected utilities associated with two choices is taken as a clear indication that one is preferred to the other. As a result, inaccuracies in an elicited utility function can cause the rejection of an action choice that is actually preferred by the decision maker.

Imprecision in the measurement of decision-maker preferences can be recognized explicitly in a decision analysis by using an efficiency criterion rather than a single-valued utility function to order alternative choices. An efficiency criterion is a preference relationship which provides a partial ordering of feasible action choices for decision makers whose preferences conform to certain rather general specifications. As such, an efficiency criterion can be used to eliminate some feasible choices from consideration without requiring detailed information about the decision maker's preferences. In many instances, the use of such a criterion may greatly reduce the number of alternatives to be considered. If enough alternatives can be eliminated, it may be possible for a final choice to be made on the basis of direct comparisons of the distributions of outcomes associated with each of the remaining alternatives.

First and second degree stochastic dominance (Hadar and Russell, 1969; Hanoch and Levy, 1969) are among the simplest and most commonly used efficiency criteria. First degree stochastic dominance holds for all decision makers who prefer more to less--i.e., for all decision makers having positive marginal utility for the performance measure being considered. Second degree stochastic dominance places an additional restriction on preferences. It requires that the decision maker's marginal utility be both positive and decreasing--i.e., it requires that the decision maker's utility function be concave. Other efficiency criteria impose additional restrictions on the decision maker's preferences or on the nature of the probability distribution of the performance measure. The mean-variance efficiency criterion (Markowitz, 1959) is simply a special case of second degree stochastic dominance in which all probability distributions are normal. Third degree stochastic dominance (Whitmore, 1970) is similar to first and second degree stochastic dominance, but it requires the additional assumption that the decision maker's utility function have a positive third derivative with respect to the performance measure.

The use of an efficiency criterion to order alternative choices is, in many respects, preferable to the use of a single-valued utility function. No direct measurements of preferences need be made. Rather, relatively easily accepted restrictions are simply imposed on the decision maker's preferences. Unfortunately, however, none of the efficiency criteria mentioned above is a particularly discriminating evaluative tool. In an application of second degree stochastic dominance by Anderson (1975), for example, 20 of 48 randomly generated farm plans were in the efficient set. Furthermore, though the restrictions on preferences required by most efficiency criteria do not appear to be unduly strict, they often run counter to empirical evidence. Again, focusing attention on second degree stochastic dominance, despite the fact that strong theoretical arguments have been made for the near universality of concave utility functions (Arrow, 1971), the weight of empirical evidence indicates that decision makers do at times exhibit increasing marginal utility (Officer and Halter, 1968; Conklin, Baquet, and Halter, 1977).

While the concept of an efficiency criterion is an attractive one, then, efficiency criteria have not always proven to be useful tools in practice. There is a need for efficiency criteria which are both more flexible and more discriminating than those described above. Furthermore, there is a need for techniques for obtaining measures of decision-maker preferences which, though less precise than those used to construct a single-valued utility function, facilitate the empirical determination of whether or not a particular efficiency criterion adequately represents the preferences of a decision maker. In the sections which follow, a more powerful efficiency criterion, stochastic dominance with respect to a function (Meyer, 1977a) is introduced, and a method for measuring decision-maker preferences designed to be used in conjunction with this criterion is presented.

Stochastic Dominance with Respect to a Function

Stochastic dominance with respect to a function is an evaluative criterion which orders uncertain action choices for classes of decision makers defined by specified lower and upper bounds, $r_1(y)$ and $r_2(y)$, on the absolute risk aversion function. The absolute risk aversion function (Arrow, 1971; Pratt, 1964), $r(y)$, is defined by the expression:

$$(5.1) \qquad r(y) = -u''(y)/u'(y)$$

where $u'(y)$ and $u''(y)$ are the first and second derivatives of a von Neumann-Morgenstern utility function $u(y)$. In the most abstract terms, values of the absolute risk aversion function are simply local measures of the degree of concavity or convexity exhibited by a decision maker's utility function. Since $u'(y)$ is assumed to be positive, a positive value of $r(y)$ implies a negative value of $u''(y)$ which in turn implies a concave utility function. Similarly, a negative value of $r(y)$ implies a convex utility function. As such, the level of absolute risk aversion also serves as a local indicator of the extent to which a decision maker is risk averse or risk loving. Following Arrow's definition, an individual is risk averse (loving) if, from a position of uncertainty, he is unwilling (willing) to take a bet which is actuarially fair (unfair).* Concavity of the utility function and risk aversion are synonymous under this definition, and both are implied by a positive value of $r(y)$. A negative value of $r(y)$ implies both local convexity of the utility function and risk loving behavior. Perhaps the most important property of the absolute risk aversion function, however, is that it is a unique measure of preferences, while a utility function is unique only to a positive linear transformation.** In effect, then, upper and lower bounds on a decision maker's absolute risk aversion function define an interval measurement in his preferences. Stochastic dominance with respect to a function orders choices on the basis of such a measurement.

More formally stated, stochastic dominance with respect to a function is a criterion which establishes necessary and sufficient conditions for the distribution of outcomes defined by the cumulative distribution function $F(y)$ to be preferred to that defined by the cumulative distribution function $G(y)$ by all agents whose absolute risk aversion functions lie everywhere between lower and upper bounds $r_1(y)$ and $r_2(y)$. As developed by Meyer (1977a), the solution procedure requires the identification of a utility function $u_0(y)$ which minimizes:

$$(5.2) \qquad \int_0^1 [G(y) - F(y)]\, u'(y)dy$$

subject to the constraint:

---

*Arrow's definition of risk aversion has been the source of some confusion, since risk aversion and risk preference have often been equated with an aversion to and a love for gambling. Unless some measure of the degree of gambling associated with a particular choice is identified as a performance measure and included as an argument in a decision maker's utility function, however, his choices are, by the omission of this factor, assumed to be unaffected by the degree of gambling involved. Arrow's concept of risk aversion refers only to the characteristics of a utility function with a single argument. As Friedman and Savage demonstrate, such a utility function can be used to explain why gambling has utility or disutility in certain situations without requiring that preferences for gambling per se be measured.

**Because a utility function is unique only to a positive linear transformation, $u(y)$ and

$$u^*(y) = a + bu(y), \quad b < 0$$

are strategically equivalent, though perhaps highly dissimilar, utility functions. The absolute risk aversion functions of these two utility functions are identical, however:

$$r(y) = -u''(y)/u'(y).$$

(5.3) $r_1(y) \leq u''(y)/u'(y) \leq r_2(y) \qquad \forall y \in [0, 1]$ *

The expression in equation (5.2) is equal to the difference between the expected utilities of outcome distributions F(y) and G(y). If, for a given class of decision makers, the minimum of this difference is positive, F(y) is unanimously preferred to G(y). If the minimum is zero, it is possible for an agent in the relevant class of decision makers to be indifferent between the two alternatives and they cannot be ordered. Should the minimum be negative, F(y) cannot be said to be unanimously preferred to G(y). In this case, the expression:

(5.4) $\int_0^1 [F(y) - G(y)]\, u'(y)dy$

must then be minimized subject to equation (5.3) to determine whether G(y) is unanimously preferred to F(y). It should be noted that a complete ordering is not ensured by the criterion. It is possible for the minimum of both equations (5.2) and (5.4) to be negative, which implies that neither distribution is unanimously preferred by the class of decision makers being considered.

Meyer uses optimal control techniques to derive the necessary and sufficient conditions for the solution of this problem. These conditions do not represent a closed form solution. Rather, they define a rule for determining the absolute risk aversion function of the utility function which minimizes equation (5.2)--a rule can be applied if the relatively unrestrictive assumption that [G(y) - F(y)] changes sign a finite number of times over the interval [0, 1] is met. Details of the solution technique are given in Meyer (1977a) and an example showing how the solution can be implemented is given in King (1979).

The major advantage of this criterion is that it imposes no restrictions on the width or shape of the relevant region of risk aversion space. The interval measurement can be as precise or imprecise as is deemed necessary for a particular decision analysis. Negative as well as positive levels of absolute risk aversion can lie within the risk aversion interval at some or all levels of the performance measure. Less flexible efficiency criteria, such as first and second degree stochastic dominance, can be viewed as special cases of this more general criterion. The requirement under first degree stochastic dominance that the decision maker have positive marginal utility places no restrictions on the decision maker's absolute risk aversion function--i.e., $r_1(y) = -\infty$ and $r_2(y) = \infty$ for all possible values of y. The requirement under second degree stochastic

---

*The range of outcomes is normalized so that all values of y fall on the bounded interval (0, 1).

dominance that marginal utility be decreasing as well as positive, on the other hand, implies that $r_1(y) = 0$ and $r_2(y) = \infty$ for all values of y.

Stochastic dominance with respect to a function is a remarkably flexible evaluative criterion which has considerable potential for use in the analysis of practical decision problems. Unlike a single-valued utility function, it does not require that an exact representation of the decision maker's preferences be specified. Unlike other efficiency criteria, it does not require that fixed restrictions be imposed on the representation of the decision makers' preferences. Furthermore, because the bounds on absolute risk aversion can be as close or as far apart as desired, stochastic dominance with respect to a function can be the basis for a more complete ordering than can be obtained with efficiency criteria such as first and second degree stochastic dominance.

An Interval Approach to the Measurement of Decision-Maker Preferences

Stochastic dominance with respect to a function is a powerful analytical tool. Before it can be used in an applied context, however, an operational procedure must be developed for the determination of lower and upper bounds on a decision maker's absolute risk aversion function. A technique for making such interval measurements of decision-maker preferences is introduced in this section. This procedure uses information revealed by a series of choices between carefully selected distributions to establish lower and upper bounds on an individual's absolute risk aversion function. The degree of precision with which preferences are measured--i.e., the size of the interval between the lower and upper bound functions--can be specified directly in accordance with the characteristics of the problem under consideration. At one extreme, the interval can be of infinite width; at the other extreme, it can converge to a single line.

The procedure for constructing interval measurements of decision-maker preferences is based on the fact that under certain conditions a choice between two outcome distributions defined over a relatively narrow range of outcome levels divides absolute risk aversion space over that range into two regions: one consistent with the choice and one inconsistent with it. The level of absolute risk aversion at which the division is made depends solely on the two distributions--i.e., their properties define the two regions. The decision maker's preferences, as revealed by his ordering of the two distributions, however, determine into which of these two regions his level of absolute risk aversion is said to fall. By confronting the decision maker with a series of choices between carefully selected pairs of distributions, the region of absolute risk aversion space which is consistent with the decision maker's preferences can repeatedly be divided. With each choice a portion of that region is shown to be inconsistent with the decision maker's preferences, and the interval measurement for the level of absolute risk aversion is narrowed. The procedure continues until a desired level of accuracy is attained. Upper and lower limits for the level of absolute risk aversion are determined at several outcome levels. These values are used to estimate upper and lower limits for the absolute risk aversion functions over the relevant range of outcome levels.

The validity of the statement that a choice between two distributions is, under certain conditions, the basis for a division of absolute risk aversion space into regions consistent and inconsistent with a decision maker's revealed preferences can be demonstrated using concepts developed by Meyer in "Second Degree Stochastic Dominance with Respect to a Function." In that paper, Meyer (1977b, p. 483) proves the following theorem:

Theorem: For cumulative distributions $F(y)$ and $G(y)$

$$\int_0^y [G(x) - F(x)]\, dk(x) \geq 0 \qquad \forall y \in [0, 1] \text{ and}$$

$$\int_0^1 [G(x) - F(x)]\, dk(x) = 0 \qquad \text{only if}$$

$$\int_y^1 [G(x) - F(x)]\, dk(x) \leq 0 \qquad \forall y \in [0, 1].$$

The theorem states that $F(y)$ is preferred to $G(y)$ by all decision makers more risk averse than the utility function $k(y)$ and that decision makers having utility function $k(y)$ are indifferent between the two distributions only if $G(y)$ is preferred to $F(y)$ by decision makers less risk averse than $k(y)$.* The function $k(y)$, then, can be considered to be a

---

*Using Pratt's definition of risk aversion in the large, a decision maker with utility function $u(y)$ is more risk averse than $k(y)$ if:

$$-\frac{K''(y)}{K'(y)} \leq -\frac{U''(y)}{U'(y)} \qquad \forall y$$

while he is less risk averse than $k(y)$ if:

$$-\frac{K''(y)}{k'(y)} \geq -\frac{U''(y)}{U'(y)} \qquad \forall y$$

Meyer (1977b) shows that $F(y)$ is preferred to $G(y)$ by all decision makers more risk averse than $k(y)$ if:

$$\int_0^y [G(x) - F(x)]\, dk(x) \geq 0 \qquad \forall y \in [0, 1]$$

and if the inequality is strict for some value of y. He also shows that $G(y)$ is preferred to $F(y)$ by all decision makers less risk averse than $k(y)$ if:

$$\int_0^1 [G(x) - F(x)]\, dk(x) \leq 0 \qquad \forall y \in [0, 1]$$

and if the inequality is strict for some value of y.

boundary function, since it separates a class of decision makers who prefer F(y) from a class who prefer G(y).

If the distributions F(y) and G(y) are defined over a narrow range of outcome levels and if the decision maker's absolute risk aversion function can be approximated by a constant value λ over that range, preference for F(y) implies that λ is greater than or equal to the minimum value of the absolute risk aversion associated with k(y). Otherwise, the decision maker would be less risk averse than k(y) and his choice would be inconsistent with expected utility maximization. Preference for G(y), on the other hand, implies that λ is less than or equal to the maximum value of the absolute risk aversion function associated with k(y), since F(y) is preferred by all decision makers more risk averse than k(y). It should be noted that the assumption that a decision maker's absolute risk aversion function can be adequately approximated by a constant value over a narrow range of outcome levels is critical here. The theorem stated above does not imply that decision makers who prefer F(y) to G(y) are more risk averse than k(y); nor does it imply that decision makers who prefer G(y) to F(y) are less risk averse than k(y). With the assumption of constant absolute risk aversion in the neighborhood of a given system output level, however, it can be inferred that decision makers who prefer F(y) to G(y) are not less risk averse than k(y) and those who prefer G(y) to F(y) are not more risk averse than k(y).

The properties of a utility function which serve as a boundary function between two distributions are dependent upon the two distributions.* By careful selection of distributions, a boundary function can be placed anywhere in risk aversion space. A series of questions can be devised, then, which allows the repeated reduction of region of risk aversion space consistent with the revealed preferences of a decision maker, thereby narrowing the interval measurement of absolute risk aversion.

A simple example will help to illustrate how the procedure works. Consider the three outcome distributions given in Figure 5.1. Each contains six possible outcomes which are said to have equal probability of occurring. Using the criterion of second degree stochastic dominance with respect to a function, it can be shown that distribution 1 is preferred to distribution 2 by all decision makers whose level of absolute risk aversion is greater than .0005 over the range of outcome levels covered by these two distributions. Distribution 2, on the other hand, can be shown to be preferred by all decision makers

*A boundary function does not exist for each pair of distributions. One would not exist, for example, if one distribution dominates the other by first degree stochastic dominance. Similarly, the existence of one boundary function does not preclude the existence of others.

1. Compare distributions 1 and 2 and indicate which one you prefer. If you prefer distribution 1, go to question 3; otherwise, go to question 2. ____

2. Compare distributions 1 and 3, and indicate which one you prefer. ____

3. Compare distributions 2 and 3, and indicate which one you prefer. ____

Distributions

1	2	3
2100	1000	1750
2400	2050	1950
2550	2650	2500
3100	3800	2750
3250	3900	3950
3450	5200	4000

Figure 5.1. A Sample Questionnaire for Interval Preference Measurement

whose level of absolute risk aversion is less than .0001. The two distributions cannot be ordered by unanimous preference over the interval (.0001, .0005), which can be termed a boundary interval in risk aversion space.* If a decision maker prefers distribution 1 to distribution 2 and if it is reasonable to assume that his absolute risk aversion function can be adequately approximated by a constant value over the range of outcome levels covered by these distributions, then it can be concluded that his level of absolute risk aversion over that range is not less than .0001, since there is unanimous preference for distribution 2 by decision makers less risk averse than .0001. Similarly, if he prefers distribution 2, it can be concluded that his level of absolute risk aversion is not greater than .0005. Preference for either one of the two distributions, then, identifies a particular portion of risk aversion space within which his own risk aversion function does not lie.

Boundary intervals can also be identified for distributions 1 and 3 and distributions 2 and 3. For distributions 1 and 3, the interval is (-.0001, .0001), with distribution 3 preferred below the boundary interval and distribution 1 preferred above it. For distributions 2 and 3, the interval is (.0005, .0010), with distribution 2 preferred below and distribution 3 preferred above.

Using this information as a guide, the series of questions at the top of Figure 5.1 was specified. They take a form similar to that of a programmed learning text. The decision maker is always asked to answer the first question, but which of the second two questions he answers will depend on the choice he makes in the first. Consider the case where the decision maker prefers distribution 2 to distribution 1 in responding to the first question. This implies that his level of absolute risk aversion is less than .0005. He is then directed to indicate his preference between distributions 1 and 3. If he prefers distribution 1, his level of absolute risk aversion is shown to be greater than -.0001. This combined with the information from the first question indicates that his level of absolute risk aversion lies on the interval (-.0001, .0005). Had he preferred distribution 3, his level of absolute risk aversion would have been shown to be less than .0001, which, when combined with the information from the first question, indicates that it lies on the interval ($-\infty$, .0001). Note that, given his response to question 1, the comparison required in question 3 would not have provided any new information. It could serve, however, as a consistency check, since preference for distribution 3 in this case would not be consistent with preference for distribution 1 in the first question.

---

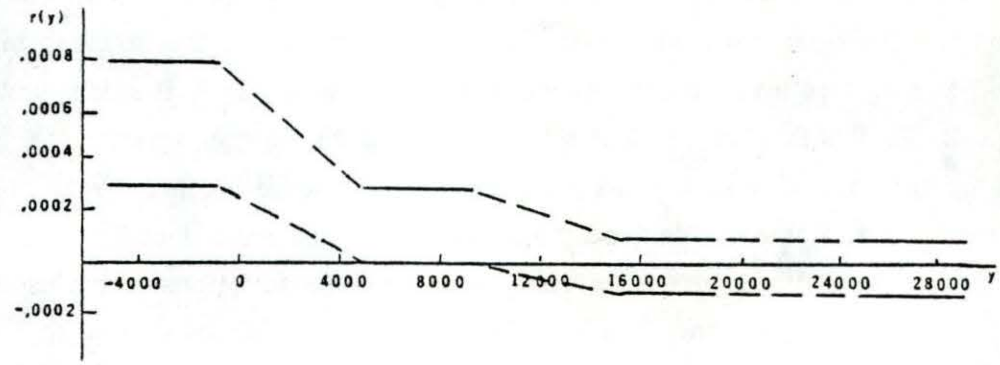
*The absolute risk aversion function associated with a boundary function for these two distributions lies everywhere within this interval over the range of outcome levels being considered.

Upper and lower bound absolute risk aversion functions constructed using this procedure for two decision makers are shown in Figure 5.2. Each is based on interval measurements made over four income ranges. Note that the slopes of the absolute risk aversion functions are not restricted. For decision maker A, the bounded interval slopes downward as income levels increase; while for decision maker B, it slopes upward and then downward. It should also be noted that the interval measurements for both decision makers contain negative as well as positive values at some income levels. When absolute risk aversion functions are derived from empirically estimated utility functions, on the other hand, their form is often severely limited by the functional form used to estimate the utility function (Lin and Chang, 1978). It should also be noted that the interval approach to the measurement of preferences avoids another common problem encountered in the estimation of single-valued utility functions. Because all questions posed require a choice between two uncertain prospects, biases due to preference for an aversion to gambling per se (Officer and Halter, 1968) are avoided. The greatest strength of the procedure, however, is its flexibility, which allows the analyst to specify the degree of precision with which preferences are measured and represented.

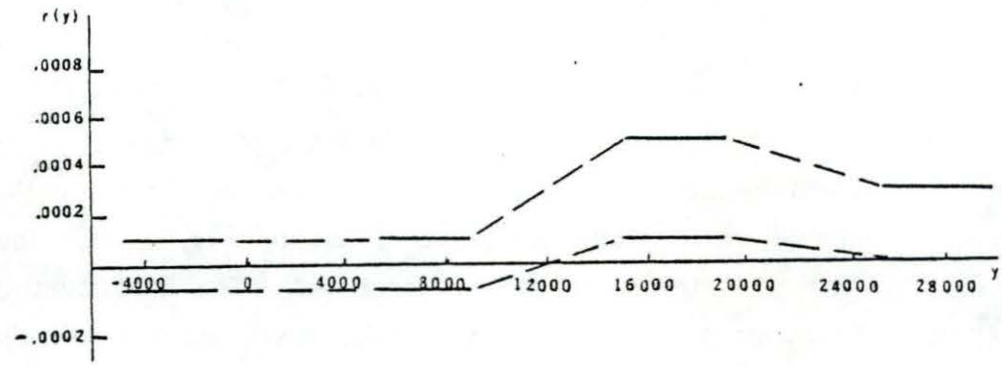
Implementation of the Procedure

The discussion in the preceding section outlines the procedure by which interval measurements of decision-maker preferences can be constructed. It does little to explain how appropriate sample distributions can be selected, however, or how the boundary interval for any pair of distributions can be identified. The implementation of the interval approach to the measurement of decision-maker preferences is described in detail by King (1979). Briefly, however, four steps in the implementation process can be identified: the specification of a measurement scale, the generation of sample distributions and the identification of a boundary interval for each pair of distributions, construction and administration of the questionnaire, and the use of interval preference measurements to order alternative choices.

Implementation of the interval approach begins with the specification of a measurement scale--a set of reference levels in absolute risk aversion space which serves as the basis for preference measurements. Because this scale determines the degree of precision with which preference measurements can be made, careful attention should be given to its specification. Two related questions which must be considered in specifying the measurement scale are: how many reference levels to include in the measurement scale and where to locate these levels on the scale.



Interval Measurement for Decision Maker A



Interval Measurement for Decision Maker B

Figure 5.2

Interval Preference Measurements for Two Decision Makers

The number of reference levels on the measurement scale depends to a large extent on the number of choices the decision maker will be asked to make in measuring absolute risk aversion in the neighborhood of any particular system output level. There is a direct relationship between the number of questions to be asked and the minimum number of reference levels on the measurement scale. In general, if a series of N choices is to be made, the measurement scale should be comprised of at least 2^N reference levels.

Once the number of reference levels to be specified has been determined, their location in risk aversion space must be established. Reference levels on the measurement scale need not be placed at regular intervals. In many instances, it is desirable to concentrate reference levels in that region of absolute risk aversion space where the decision maker's actual level is expected to fall or in the regions where relatively small changes in absolute risk aversion have the greatest impact on preference orderings. Experience to date indicates that most of the detail on the measurement scale should be concentrated in the risk aversion interval between -.0001 and .0010. Actual measurements for a variety of decision makers have tended to fall most frequently within this interval, and tests on several empirical decision problems have indicated that choices are most strongly affected by changes in absolute risk aversion within this range.

Once a measurement scale has been specified, sample probability distributions which serve as the basis for the choices used to reveal the decision maker's preferences must be generated, and a boundary interval must be identified for each pair of distributions. These tasks are accomplished by a computer program for which the measurement scale serves as an input.

The sample distributions are constructed in a random manner by generating several hundred variates from a user-specified distribution and grouping them into sets of six observations each. Each set of observations is considered to be a distribution of outcomes, and each element is said to have a one-sixth probability of occurrence. These distributions should be defined over a relatively narrow range of outcome levels, since the decision maker's level of absolute risk aversion is said to be constant over that range.* In past applications, six elements have been included in each distribution. More complex distributions would make decision makers' choices unduly difficult, while distributions with fewer elements may not be rich enough to make the choice interesting. The use of six-element distributions also facilitates explanation of the choice situation to the

---

*Experience to date indicates that a range of five to ten percent of the entire range of system output levels over which preferences are to be measured is adequate.

decision maker, since the probability of any one element occurring can be equated directly to the probability of obtaining a specified number of dots on a single role of a die. There is no theoretical reason why distributions with either a larger or smaller number of elements cannot be employed, however.

Given the measurement scale, the program that generates the sample distribution uses the criterion of second degree stochastic dominance with respect to a function to identify the narrowest boundary interval for each pair of distributions. Absolute risk aversion intervals defined by adjacent reference levels on the measurement scale are tested iteratively to identify (1) the highest reference level, λ_1, such that all decision makers less risk averse than λ_1 prefer one distribution, and (2) the lowest reference level, λ_2, such that all decision makers more risk averse than λ_2 prefer the other distribution. These two reference levels define the boundary interval (λ_1, λ_2) for that pair of distributions.

At least one pair of distributions for which the boundary interval lies between any two adjacent levels on the measurement scale should be identified by this procedure if enough sample distributions are considered. Once this has been done, a series of questions can be formulated, with each question focusing on a different interval. As was noted earlier, the questions take a form similar to that of a programmed learning text, directing the decision maker through a hierarchy of comparisons designed to continually increase the precision of the interval measurement. In general, each question should focus on an interval in the center of the region of absolute risk aversion space consistent with the decision maker's prior choices. The number of outcome levels at which direct measurements of absolute risk aversion are to be made is determined by the analyst. Experience to date has shown that direct measurements in the neighborhood of three to four outcome levels provide an adequate basis for the construction of an absolute risk aversion function over even a broad range of outcomes. If, for example, annual income is the performance measure for which preference information is to be elicited and the relevant income range is from 0 to \$20,000, direct measurements of absolute risk aversion could be made in the neighborhood of \$3,000; \$10,000; and \$17,000.

Before the questionnaire is administered, the decision maker should have a clear understanding of its objective, which is to obtain an accurate representation of his preferences. The outcome measure for which preferences are to be measured should already have been clearly defined and should be recognized by the decision maker to be the primary performance indicator he will consider when making a choice in the situation being analyzed.

Administration of the questionnaire and interpretation of the results are straightforward. Completion of a questionnaire comprised of four three-question series takes approximately 20 minutes. Experience to date has shown that decision makers find this preference elicitation procedure more interesting and more informative than the interview process required to elicit a single-valued utility function.

Finally, interval preference measurements are used in combination with the evaluative criterion of stochastic dominance with respect to a function to order the action choices being considered by a decision maker. Upper and lower bound absolute risk aversion levels for each of the outcome ranges over which direct preference measurements have been made and the outcome distributions to be ordered must be specified by the user of the computer program which accomplishes this task.

In many decision situations, a large, if not infinite, range of choices may be open to the decision maker. Therefore, a systematic technique for the identification and evaluation of a large number of possible strategies is also needed if interval preference measurements and the criterion of stochastic dominance with respect to a function are to be of use in the solution of practical decision problems. The generalized risk efficient Monte Carlo programming model developed by King incorporates the criterion of stochastic dominance with respect to a function into an optimization framework. This procedure, which is described in detail elsewhere (King, 1979), is flexible enough to be applicable in a wide range of decision situations and does not require that important simplifying assumptions be made concerning the representation of decision-maker preferences, the form of the outcome distributions associated with the strategies considered, or the nature of the decision problem itself.

An Empirical Test

A simple experiment was designed and conducted to test the efficacy of the interval approach to the measurement of decision-maker preferences. Three questionnaires were administered to a group of graduate research assistants in the Department of Agricultural Economics at Michigan State University. The first questionnaire employed the procedure described in the preceding sections to obtain an interval measurement of each subject's absolute risk aversion function. The second questionnaire was used to elicit information required for the construction of a single-valued utility function for each subject. Finally, in the third questionnaire, the respondents were asked to make a series of six choices between pairs of distributions, each distribution being comprised of six elements and each being defined on the interval over which preferences had been measured. Information from the first two questionnaires was used to predict the choices made by each respondent in the third questionnaire, and these predictions were compared to the actual responses. In this way, the accuracy of each of the two approaches to the measurement of preferences was tested.

In evaluating each approach, two criteria were considered: the number of correct predictions and the number of choices for which a definite ordering was made. A prediction was said to be correct if the respondent's actual choice was not excluded from the efficient set of choices and incorrect if it was excluded. The preference measure having the highest proportion of correct predictions was said to be the more accurate according to this criterion. Concern with the proportion of correct predictions is analogous to concern with the probability of Type I error in a statistical test, the latter being the probability that a true statement will be judged to be false and be rejected. This measure of accuracy is not a good indicator of the relative discriminatory power of preference measurements based on these two approaches. The criterion of first degree stochastic dominance, which holds for all decision makers who prefer more of the performance measure to less, should never exclude a preferred choice from the efficient set and so should be perfectly accurate according to the criterion defined above. Often, however, it also fails to exclude many choices from the efficient set. A single-valued utility function, on the other hand, is the basis for a complete ordering of choices--i.e., it always leads to an efficient set having a single element. Therefore, the number of choices actually ordered was also considered. Concern with this measure of discriminatory power is analogous to concern with the Type II error associated with a statistical test, which is the probability that a false statement will be judged to be true and not rejected.

Clearly there are trade-offs between the accuracy and the discriminatory power of a preference measurement. Unlike other measurement techniques and evaluative criteria, the combined use of interval preference measurements and stochastic dominance with respect to a function permits explicit consideration of these trade-offs. As the precision of the interval measurement increases, it becomes a more discriminating basis for the ordering of choices, but the probability of excluding preferred choices from the efficient set also increases. Such trade-offs between accuracy and discriminatory power were also analyzed in the experimental test of the interval approach to the measurement of preferences. Direct interval measurements of absolute risk aversion were made at three levels of income--the relevant performance measure in this instance. These measurements were based on a sequence of four questions at each income level. By constructing interval measurements on the basis of information available at the end of each question, however, four preference measurements--each more precise than the one which preceded it--were made for each subject.

Nine of ten subjects correctly completed all three questionnaires. Since each subject made six choices on the third questionnaire, each preference measurement was used to predict a total of fifty-four choices. The results of the experiment are presented in Table 5.1. They show that there is a clear trade-off between accuracy and discriminatory power. First degree stochastic dominance and the single-valued utility function are at opposite extremes in this trade-off relationship, and the interval measurements are arrayed between the two. Several factors should be noted. With regard to the accuracy of the interval measurements, it falls at a relatively constant rate as the number of questions posed increases, but even at the higher levels of precision it exceeds that realized with the single-valued utility function. The discriminatory power of the interval measurements, on the other hand, increases dramatically as the number of questions asked at each income level increases. In contrast, first and second degree stochastic dominance clearly do not discriminate well among the distributions which were the basis for the decision makers' choices.

Concluding Remarks

The interval approach to the measurement of decision-maker preferences and the criterion of stochastic dominance with respect to a function greatly facilitate the application of decision theory based on the expected utility hypothesis in the analysis of practical decisions. As was noted earlier, they can be used in conjunction with a general procedure for the identification of preferred choices to further extend the usefulness of this body of theory in the solution of complex decision problems. These techniques need further refinement, however.

Referring specifically to the interval preference measurement technique, which was the focus of this paper, there is a need for further experiments to identify measurement scales that will be well suited for use in a variety of decision situations and a need for improvements in the mode of questioning by which information on preferences is elicited. Research which will be useful in making such refinements is currently underway at Michigan State University. Another important need is for more research on the representation of preference relationships which depend on more than one performance criterion and for the development of multivariate stochastic dominance criteria. Some work has been done in the latter area by Levy and Paroush (1974) and by Kihlstrom and Mirmon (1974), but further research is needed. Particularly valuable would be an extension of stochastic dominance with respect to a function to the multivariate case.

Table 5.1. Performance Indicators for Alternative Preference Measures[a]

Performance Indicator	Interval Measurement Number of Questions				Single-Valued Utility Function	First Degree Stochastic Dominance	Second Degree Stochastic Dominance
	1	2	3	4			
1. Percent of choices predicted correctly[b]	98	88	78	72	65	100	98
2. Percent of choices ordered	9	50	83	91	100	0	7

[a]We thank Garth Carmen, who helped to conduct this experiment.

[b]A choice was said to be predicted correctly if the preferred distribution was not excluded from the efficient set.

CHAPTER VI

AN EMPIRICAL MODEL FOR ENERGY SUPPLY INVESTMENT AND DISINVESTMENT DECISION MAKING UNDER UNCERTAINTY

Michael H. Abkin and Gary R. Ingvaldson

Introduction

Investment/disinvestment theory describes the economic decisions associated with production processes using as inputs the services of durables as well as non-durables. Specificially, it considers simultaneously questions of (1) when to acquire (invest in), or (2) salvage (disinvest in) durable assets, (3) what size of a durable to acquire, (4) how intensively to extract services from a durable, and (5) how much to maintain a durable. Decision theory, meanwhile, attempts to describe the process whereby decision makers choose among alternative courses of action (or "strategies") whose outcomes are uncertain because they depend on conditions in the environment (or "states of nature") which are themselves uncertain. Chapter II of this report presents a review of decision making in electric utilities with respect to investments and disinvestments under conditions of uncertainty, and Chapter V discusses practical approaches to implementig the theory of decision making under uncertainty. These two bodies of theory are integrated and extended in Chapter IV, and it is the objective of this chapter to present a model and computer software package which implements this integrated theory for practical application in decisions of investment and disinvestment in the energy supply industry.

The terms of our EPRI contract refer to analysis of the impacts of asset fixity and uncertainty on energy supply in general. However, in order to have a well-defined problem, we found it necessary to focus on firm-level decisions involving increases and decreases in electric power generating capacity. Nevertheless, the theory developed is general to any production process involving durable factors, and it would be straightforward to generalize the model presented here to address other classes of energy supply decisions, such as oil or gas drilling, coal mining, pipeline distribution, electric power transmission, etc. Indeed, the software package developed for this model (Abkin and Ingvaldson, 1980) is suitable for any such application, reserving the model specifications particular to that application to user-supplied sub-routines. For illustrative purposes, however, the remainder of this discussion and the general test case presented as an example refer to decisions with respect to electric power generating capacity.

The next section casts the decision problem in an optimal control framework and suggests the feasibility of numerically obtaining a global solution in the deterministic case. Because of the uncertainty, however, this problem is decidedly not deterministic.

Therefore, succeeding sections (1) describe the generalized risk-efficient, Monte Carlo programming (GREMP) approach for handling the stochastic case, and (2) discuss the assumptions and results of a general test case designed to illustrate the application of the GREMP approach to this problem. Chapter VII compares this model with the minimum revenue requirements method of investment decision analysis in the context of the 1977 Boardman River hydro decision problem in Grand Traverse County, Michigan. The concluding section of the present chapter summarizes recommendations for further developments and refinements of the model.

The Optimal Control Framework

The general investment/disinvestment problem is a dynamic one involving optimal decisions with respect to when durable factors of production are to be acquired and in what sizes, when they are to be disposed of, at what rate services are to be extracted from them during their lifetime, and when and how much they are to be maintained to extend that lifetime. An electric utility is a single firm which typically operates several generating plants, each contributing to meeting the firm's load demand. These plants, or individual generators within them, are the durables for which the utility must make the above decisions (as well as other durables, of course, such as transmission lines, etc.). Since the utility generally dispatches plants according to their relative efficiency and is required to meet its load and maintain a reserve capacity, these investment/disinvestment/utilization/maintenance decisions cannot be made for any one plant independently of the others. Therefore, the general problem is stated in an optimal control framework in order to maintain the dynamic and simultaneous character of the decisions to be made in the context of the integrated system.

Two caveats, as obvious as they are, must nevertheless be noted with respect to the optimal control problem stated below. First, there is no implication that a one-time solution to the problem will dictate the pattern and schedule of investment/disinvestment/utilization/maintenance activities over the next T years of the planning horizon. In practice, and even in theory, such an implication would be patently ridiculous. Rather, if the model were to be used in practice, it would typically be solved repeatedly at intervals--most likely frequent intervals--dictated by the acquisition of new information on exogenous variables (e.g., prices, regulations, and load forecasts) and control options (e.g., generation technologies).

Secondly, in any case, this model is primarily an economic model, and practical decisions typically require a great deal of additional, non-economic information--technical, political, social, legal, ecological, etc. Therefore, the model presented here can never claim or be expected to "dictate" decisions; it can at best be only one, hopefully credible source of information for decision making.

The general optimal control problem is stated as follows:

given a system represented by the state equations:

$x(t) = a[x(t), u(t), z(t)]$

where x = vector of state variables

u = vector of control variables

z = vector of noncontrollable exogenous variables

and where $x(t)$ is required to meet certain constraints at every point in time, i.e., $x(t) \in X$ for all $t \in [0,T]$, where X is the set of admissible states and T is the planning horizon;

find a control history $u(t)$, also subject to constraints $u(t) \in U$ for all $t \in [0,T]$, where U is the set of admissible controls, in order to

maximize the objective function $J = h[x(T)] + \int_0^T g[x(t), u(t), z(t)]\, dt$

where h is the contribution to the objective of the state of the system at the end of the planning period and g is the accumulation of contributions over time.

Variable Definitions

a) Dimensions, n and k

For an electric utility operating a number of generating plants, assume k is the number of plants in existence at time zero in the analysis, and n is the maximum number of plants to be considered in the analysis. Therefore, n-k is the maximum number of additional plants to be considered as options for replacing or augmenting existing capacity over the planning horizon T. Each of the n plants, existing as well as potential, is given a set of characteristics at the beginning of the analysis, e.g., size, operating efficiency, and fuel type (coal, oil, nuclear, etc.). Therefore, in order to consider a number of options, n-k will typically be larger than the total number of plants that (or represent greater capacity than) is likely to be needed T years in the future. Note that each of the n "plants" may be interpreted, depending on the needs of the analyst, as a plant, as a generator, or as an aggregation of plants of similar vintage and characteristics.

b) State Variables: $x_i(t)$, $i = 1, 2, ..., 3n$

There are 3n state variables. That is, for each plant j, j=1, 2, ..., n:

$GCAP_j(t)$ = plant capacity (megawatts)

$VALDUR_j(t)$ = unit market value of the durable ($/megawatt)

$VALSER_j(t)$ = value of services to date ($)

c) Control Variables: $u_i(t)$, i = 1, 2, ..., 6n-2k+1

There are 6n-2k+1 control variables, some applicable to all n plants and others applicable only to the n-k plants not in existence at time zero.

For each plant j, j = 1, 2, ..., n:

$UTILT_j(t)$	=	time utilization rate (proportion of total hours per year)
$UTILC_j(t)$	=	capacity utilization rate (proportion of megawatt capacity)
$VMAINT_j(t)$	=	variable maintenance rate to replace lost or used capacity (megawatts/year)
$STIME_j$	=	time to salvage or disinvestment (year)

For each plant j not existing initially, j = k+1, k+2, ..., n:

$CAPO_j$	=	initial plant capacity (megawatts)
$ACQT_j$	=	time of acquisition or investment (year)

Finally, there is one control variable which is not plant specific:

PINRG (t)	=	energy purchased or (if negative) sold (megawatt-hours/year)

d) Exogenous Variables: $z_i(t)$, i = 1, 2, ..., 8n+2

There are 8n+2 exogenous variables which affect the performance of the system but which are assumed to be either beyond the control of the firm, or at least assumed given for purposes of this analysis. These are known functions of time. Included are prices, costs, discount and depreciation rates, and load forecasts.

Specifically, for each plant j, j = 1, 2, ..., n:

$FOMC_j(t)$	=	fixed operating and maintenance costs ($/megawatt-year)
$SCHED_j$	=	scheduled down time for regular maintenance (proportion of hours/year)
$VMNTC_j(t)$	=	variable maintenance cost ($/megawatt)
$VICST_j(t)$	=	cost of variable input (fuel) ($ per unit of fuel, e.g., barrel, ton, etc.)
$ACCST_j(t)$	=	plant acquisition cost ($/megawatt); this is also a function of the size of the plant, $CAPO_j$
$DCOST_j(t)$	=	relicensing cost ($/megawatt); depends on time from acquisition
$D_j(t-ACQT_j)$	=	depreciation schedule (called "time cost" in Chapter IV) (proportion per year); depends on time from acquisition

In addition, there are two variables common to the firm as a whole:

XLOAD(t) = consumer demand (megawatt-hours per year)

R(t) = opportunity cost of capital (called "control cost" in Chapter IV) (proportion/year)

Note the simplifying assumption that load is neither price responsive nor controllable by the firm. This assumption could be relaxed, if desired, to include load management policies in the analysis and to capture the secondary impacts of the firm's investment and operating decisions on its load through the primary impacts on price.

Model Equations

This section describes the state equations representing the dynamics of the system's behavior, the calculation of associated performance variables, the constraints imposed on state and control variables, and the objective function to be optimized. First, for each plant j, j = 1, 2, ..., n, the power output capacity, GCAP, changes over time according to the following differential equation:

$$(6.1) \qquad \frac{d}{dt} GCAP_j(t) = \begin{cases} VMAINT_j - USCST_j(t) & \text{for } ACQT_j \leq t \leq STIME_J \\ 0 & \text{else} \end{cases}$$

with boundary conditions

$$GCAP_j(t) = \begin{cases} CAPO_j & \text{for } t = ACQT_j \\ 0 & \text{for } t < ACQT_j \text{ and } t > STIME_j \end{cases}$$

where for plants existing at time zero (j = 1, 2, ..., k) we can assume $ACQT_j = 0$, and where USCST is the rate at which power output capacity is lost, in megawatts/year, due to intensity of utilization.

In general, this capacity user cost can be defined as a function of plant capacity, of time utilitization, and of capacity utilization:

$$(6.2) \qquad USCST_j(t) = c_j\,[GCAP_j(t),\ UTILT_j(t),\ UTILC_j(t)]$$

Such loss in capacity is an important concept in this model, for it influences forced outages and the associated variable maintenance necessary to replace it.

Variable maintenance VMAINT, remember, is a control variable determined in the optimum solution. The analyst may want to specify VMAINT in a control law, such as:

$$(6.3a) \qquad VMAINT_j(t) = m_{1j}\,[CAPO_j - GCAP_j(t)]$$

or

$$(6.3b) \qquad VMAINT_j(t) = m_{2j}(USCST_j)$$

or perhaps some combination of these. In such a case, optimization would be over the parameters of the control law or over alternative control laws.

The unit market value of the durable (the generating plant) changes over time due to such market conditions as the introduction of improved technologies, changes in the relative prices of fuels, labor and other inputs, and changes in market institutions--the pure time cost discussed in Chapter IV. This is captured in the model by an exogenously specified depreciation schedule, as follows:

$$(6.4) \qquad \frac{d}{dt}\, VALDUR_j(t) = \begin{cases} -D_j(t-ACQT_j)\; VALDUR_j(t) & \text{for } ACQT_j \le t \le STIME_j \\ 0 & \text{else} \end{cases}$$

with boundary conditions:

$$VALDUR_j(t) = \begin{cases} ACCST_j(t) & \text{for } t = ACQT_j \\ 0 & \text{for } t < ACQT_j \text{ and } t > STIME_j \end{cases}$$

where D_j depends, in general, on the time from acquisition of plant j.

Salvage value, in dollars, then is the market value of the plant less a "decommissioning" cost representing disposal costs in general:

$$(6.5) \qquad SALVAL_j(t) = VALDUR_j(t)\; GCAP_j(t) - DCOST_j(t)\; CAPO_j$$

Notice that the disposal costs are based on the initial size of the plant rather than its current effective capacity. Thus, salvage value is reduced by time cost through VALDUR [equation (6.4)], and increased by maintenance and reduced by user cost through GCAP [equation (6.1)].

The value of services, VALSER, generated by each plant is an important state variable, because it determines the economic life of the plant and whether the plant should be acquired in the first place, as discussed in Chapter IV. Since VALSER is derived from the objective function J of the optimal control problem, we turn our attention now to the latter before looking at the former.

For investment/disinvestment decisions, the appropriate objective is to maximize the discounted present value of economic gains (in \$) accumulated over a suitable planning horizon [0, T]. Therefore, for the objective function J given above, we define h = 0 and g as the discounted gain function (in \$/year). A gain function is used rather than a profit function in order to account for the additional time, control, and user costs (Chapter IV) associated with durable factors of production.

Therefore, for discount rate ρ:

$$\begin{aligned}(6.6) \qquad g(x,u,z) = e^{-\rho t}\bar{g}(t) = e^{-\rho t}\Big\{ & [CPR(t) - PPR(t)]\; PINRG(t) \\ & + \sum_{j=1}^{n} [HPY \cdot CPR(t)\; U_j(t)\; GCAP_j(t) - VICST_j(t)\; VIUSE_j(t) \\ & + [VALDUR_j(t) - VMNTC_j(t)]\; VMAINT_j(t) - R(t)\; SALVAL_j(t) \\ & - D_j(t)\; VALDUR_j(t)\; GCAP_j(t) - VALDUR_j(t)\; USCST_j(t) \\ & - FOMC_j(t)\; CAPO_j - RCOST_j(t) - TAX_j(t)] \Big\}\end{aligned}$$

where HPY = 8760 hours/year, VIUSE = variable input (e.g., fuel) used by the plant (units of fuel/year), CPR and PPR are the consumer price of energy and the price to the firm of purchased energy, respectively (both in \$/megawatt-hour), and where the total plant utilization, U_j, is:

$$(6.7) \qquad U_j(t) = UTILC_j(t)\ UTILT_j(t)$$

Note that the control cost, i.e., the capital cost of the plant, $R(t)\ SALVAL_j(t)$, is included in the objective function.

The value of services for each plant, then, is defined as that part of the gain which is attributable to the services extracted from that plant. Following the theorem presented in Chapter IV, it is determined by the following differential equation, discounting to the plant's acquisition time:

$$(6.8) \qquad \frac{d}{dt} VALSER_j(t) = \begin{cases} e^{-\rho(t-ACQT_j)} \displaystyle\int_0^{U_j(t)} \frac{\partial \bar{q}(t)}{\partial u_j}\, du_j & \text{for } ACQT_j \le t \le STIME_j \\ 0 & \text{else} \end{cases}$$

with the boundary condition:

$$VALSER_j(t) = 0 \text{ for } t < ACQT_j$$

This must be done holding the use of other fixed and variable inputs in a fixed relationship to the durable's use. In the case of one durable j, such as an electric power plant, and one variable input, such as fuel, such a relationship could be, for example:

$$(6.9) \qquad VIUSE_j(t) = HPY \frac{HEATR_j}{FUELC_j} U_j(t)\ GCAP_j(t)$$

where HEATR = the heat rate of the plant (BTU/megawatt-hour)

FUELC = the heat content of the plant's fuel (BTU/unit of fuel)

Defining the net price:

$$NP(t) = CPR(t) - VICST_j(t) \frac{HEATR_j}{FUELC_j}$$

and referring to equations (6.2) and (6.3), we can derive:

$$(6.10) \qquad \frac{\partial \bar{q}(t)}{\partial U_j} = HPY \cdot NP_j(t) GCAP_j(t) - VMNTC_j(t) \frac{\partial m_j}{\partial U_j}$$

$$+ [HPY \cdot NP_j(t) U_j(t) - (D_j(t) + R(t)) VALDUR_j(t)] \cdot [\int_0^t \frac{\partial m_j}{\partial U_j} d\tau - \int_0^t \frac{\partial c_j}{\partial U_j} d\tau]$$

which, given explicit functions for user cost and maintenance (c_j and m_j, respectively), enables us to find $VALSER_j$ by equation (6.8). If a control law is not specified for maintenance, then $\partial m_j / \partial U_j = 0$.

As discussed below, further theoretical development is necessary to handle the case of multiple, mutually dependent durables. Therefore, the empirical model implemented does not attempt to isolate the portion of gain attributable to each durable itself from that of other durable and variable inputs. Thus, for the time being, the total gain generated by the operation of a durable [i.e., for each plant j under the summation sign of equation (6.6)] is used in place of the integral in equation (6.8) in computing $VALSER_j$.

Knowing the value of services for each plant enables us to use the control law for salvage (or replacement) time derived in Chapter IV. Assuming upon salvage a plant is replaced with one identical to it, STIME is determined by comparing the current (not discounted) rate of change of VALSER with the annuity value (annualized average) of VALSER. The annuity value of the existing plant is used as a surrogate for that of the identical replacement, implicitly assuming the replacement experiences the same history as the original plant. Thus:

$$(6.11) \qquad STIME_j = t:\ e^{\rho(t - ACQT_j)} \frac{d}{dt} VALSER_j(t) \leq \frac{\rho}{1-e^{-\rho(t-ACQT_j)}} VALSER_j(t).$$

Other optimal life criteria could be derived based on other assumptions made regarding the nature of the replacement, e.g., salvage without replacement, technological change, etc. In any case, as in the identification of value of services, the theory upon which equation (6.11) and any such alternative derivations are based (Chapter IV) does not yet satisfactorily capture the situation of multiple, mutually dependent durables.

The consumer price, CPR, of the firm's product (e.g., electric power) may be specified as a function of other variables in the model, or it may be projected independently and input to the model. In the latter case, CPR would be included in the list of exogenous variables, z_i, given above.

In the case of a regulated electric utility, the price to the customer includes a pass-through of costs (fuel, maintenance, purchased power, depreciation, etc.) plus an allowed rate of return on the capital investment. Referring to the objective function [equation (6.6)], since costs are offset on the revenue side by CPR, and insofar as the allowed rate of return included in CPR compensates for the control costs (ignoring time lags in regulatory adjustments), we can see that the objective function will tend to be approximately zero for whatever control strategies are implemented. That is, the regulated price will automatically adjust to maintain a zero economic gain.

In such a situation, the objective function specified in equation (6.6) is unable to distinguish among control strategies. Therefore, the empirical model implemented here uses, for electric utilities regulated in this way, an objective function which omits the revenue side and accounts only for the costs--essentially a cost minimization problem. That is, the implied objective is to minimize the discounted costs to the consumer.

Another implication of this type of price regulation is that the value of services generated over the lives of the firm's plants will average out to zero. Plants of above average efficiency will have positive value, and those of below average efficiency will have a negative value of services.

An important constraint on the system (for an electric utility) requires that the load must be met at every point in time, i.e.,

$$(6.12) \qquad PINRG(t) + \sum_{j=1}^{n} HPY \cdot U_j(t)\, GCAP_j(t) = XLOAD(t).$$

Other constraints on state and control variables are:

$$(6.13) \qquad 0 \leq GCAP_j(t) \leq CAPO_j \qquad \text{for } j = 1, 2, \ldots, n$$

$$(6.14) \qquad CAPO_j \in F_j \qquad \text{for } j = k+1, k+2, \ldots, n$$

$$(6.15) \qquad 0 \leq ACQT_j \qquad \text{for } j = k+1, k+2, \ldots, n$$

$$(6.16) \qquad ACQT_j \leq STIME_j \qquad \text{for } j = 1, 2, \ldots, n$$

$$(6.17) \qquad UTILC_j(t) \in \{0, [UCMIN_j, 1]\} \qquad \text{for } j = 1, 2, \ldots, n$$

$$(6.18) \qquad 0 \leq UTILT_j(t) \leq 1\text{-}SCHED_j\text{-}FORC_j(t) \qquad \text{for } j = 1, 2, \ldots, n$$

Note that equation (6.14) says the initial plant capacity may be constrained to selected sizes or a certain range of sizes F; and equation (6.17) says, if a plant is going to be operated at all at time t, it must be utilized at least UCMIN percent of capacity. Equation (6.18) is necessary to allow for scheduled and forced down time.

Approaches to Finding the Optimal Solution

The size and degree of non-linearity of this problem make an analytic solution virtually impossible. However, there are a number of techniques available, some in "canned" software packages, for finding numerical solutions (e.g., see Kirk, 1970, and Luenberger, 1973). The gradient projection method and adaptations of it appear particularly suitable for our constrained problem; indeed, this method requires the control variables to be in a bounded region.

We have not gone this route, because of the part of the problem we haven't discussed yet--the uncertainty part. Uncertainty is introduced by specifying probability density functions for a subset, if not all, of the above exogenous variables, as we have done for the illustrative test case described below. While a numerical solution appears to be feasible (although probably costly) for the deterministic case, a different approach must be found for the stochastic case. This is not to say that finding a numerical, deterministic solution to the problem described above wouldn't be instructive and useful--indeed, it certainly would be--but we have not done so in this project. For the stochastic case, King (Chapter V) suggests an approach which is used here as described in the next section.

GREMP for the Stochastic Case

We have elected to solve for the stochastic case with the GREMP package as presented in King (1979). GREMP is an acronym for "generalized risk efficient Monte Carlo programming." The technique is particularly well suited for problems in which it is difficult or impossible to determine a solution analytically. It does not necessarily identify the optimal strategy but rather a nearly optimal strategy. This is accomplished by examining a large number of alternatives under a variety of states of nature and selecting those strategies which perform "best" according to a given criterion.

Three major processes are included in the model--strategy generation, simulation and distribution of outputs, and evaluation. These are illustrated by the flow chart in Figure 6.1. Strategy generation may be accomplished through (1) a process of random selection, (2) specification of a set of strategy choices deemed apt by the decision maker, (3) an experimental design, or some combination of these. Each strategy is simulated repeatedly for a number of states of nature, and a distribution of outputs likely for that strategy is generated. The distributions of outputs for the various strategies are then evaluated by applying the criterion of stochastic dominance with respect to a function. Those strategies which are not dominated by any other comprise the efficient set of strategies.

Designation of Strategies

King, in his application of the GREMP model, constructed strategies at random. We have chosen instead to furnish a pre-selected set of strategies to the model. Random selection would have involved choices of plant size, plant type, acquisition time, utilization rate, and other variables for a number of plants. The combination of all these factors would have yielded a vast magnitude of strategies, many of which would obviously be inappropriate. Computational constraints would limit random evaluation to a very small fraction of the total, resulting in possible omission of favorable strategies.

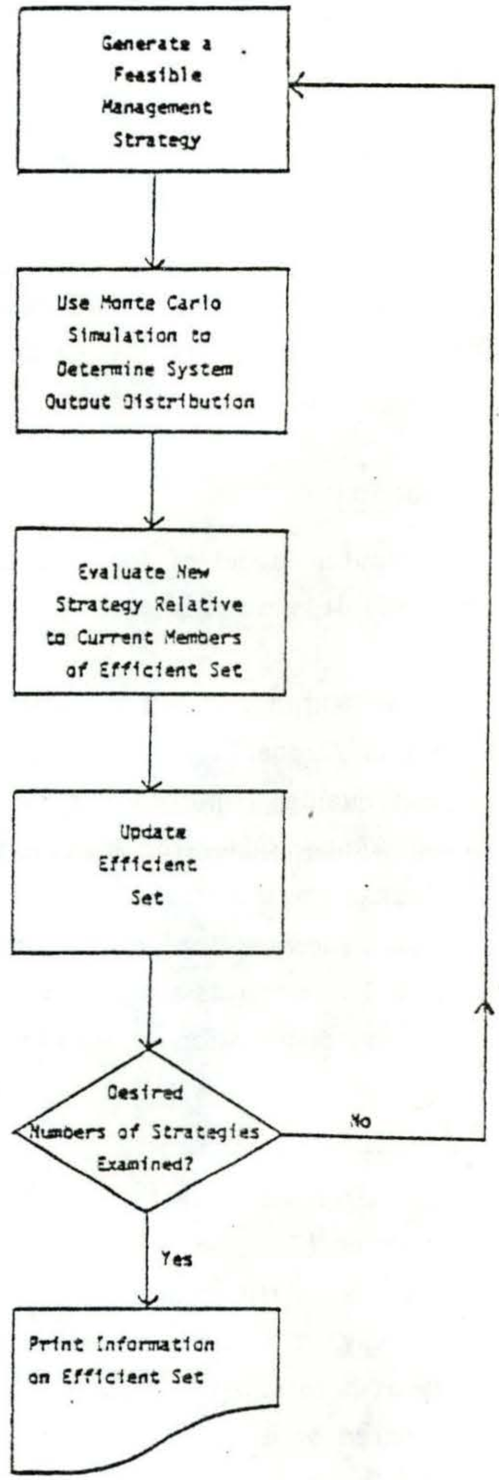


Figure 6.1

A Flow Chart of the GREMP Model

Source: Figure taken from King (1979).

For the electricity generating capacity decision problem, an experimental design has been constructed, for the illustrative test case described below, for plant capacities and acquisition times, while decision rules have been implemented for utilization rates and maintenance and replacement policies. The a priori selection of the experimental design enables the analyst to tap the expertise of the decision maker and leads to a more efficient use of computing resources. Similarly, standard decision rules are useful for such variables as the utilization rate of a plant. It is dependent on the load and the relative efficiencies, capacities, and utilization rates of itself and other plants and, thus, is better specified in the form of a decision rule (i.e., a dispatch rule) also drawn from knowledgeable expertise, as opposed to a pre-specified rate.

Simulation and Distribution of Outputs

We have constructed a computer model of the investment/disinvestment problem presented in the previous section. It simulates each strategy under the same set of random states of nature.

The exogenous variables represented within the states of nature include consumer demand and prices of variable inputs. Probability density functions are furnished for each of these variables and a random number generator is used to construct the states of nature. The seed of the random number generator is reset for each strategy to ensure that all strategies are simulated under the same set of conditions.

Given a strategy, the model generates the gain (actually, cost) function for that strategy for each state of nature. These results are then ordered from smallest to largest gain in order to obtain the cumulative distribution of gains for that strategy.

Stochastic Dominance with Respect to a Function

The GREMP model evaluates alternative strategies under the criterion of stochastic dominance with respect to a function (Chapter V). Each strategy is evaluated after its cumulative distribution of gains has been generated. It is compared with other alternatives and if dominated by any, it is removed from further consideration. If not dominated by any other strategy, it is retained for further consideration. Any previously retained strategy which is dominated by a new strategy is dropped at this time. Those strategies remaining after all have been evaluated are referred to as the efficient set of strategies. Within the criterion specified, it is not possible to proclaim any one to be superior to the others.

Stochastic dominance with respect to a function does not necessarily yield a unique strategy. Rather, it provides the decision maker with a set of strategies favorable to maximizing the gain function under uncertain conditions. The efficient set may not include a global optimum but a large sample of wisely selected alternative strategies assures that nearly optimal strategies will be identified and gives a high probability that the global optimum is included.

A General Test Case

In order to fully test and illustrate the decision features of the model, as an operationalization of the investment/disinvestment/utilization/maintenance decision theory, we have developed a general test case of an electric utility with a number of existing generating plants, a number of options for future plants, and a load forecast. A specific application to the Boardman River hydroelectric sites in Grand Traverse County is discussed in Chapter VII. There is much that could be done to make this general test case more realistic; therefore, the assumptions and results presented in this chapter must be considered tentative and for illustrative purposes only.

Assumptions

The assumptions defining the general test case (itemized below) include explicit functions and parameter values for the model presented above and in some cases variances from that model.

1. Planning Horizon and Discount Rate

$T = 40$ years, $\rho = 11.75\%$

2. Plants To Be Considered

Four plants are assumed to exist at time zero. The number of plants to be analyzed for expansion of capacity will vary with the strategy, but four different sizes and types of such additional plants are assumed to be available. Table 6.1 summarizes the characteristics of the initial and additional plants.

3. User Cost, Forced Outage, and Variable Maintenance

The explicit functions for user cost [c_j in equation (6.2)], variable maintenance [m_j in equation (6.3)], and forced outage for a plant j are assumed to derive from the degree to which the plant's capacity utilization exceeds a threshold level and the duration of that excess.

Table 6.1

Characteristics of the General Test Case
Electric Power Generating Plants

Characteristics	Initial Plants				Additional Plants			
	1	2	3	4	A	B	C	D
Fuel Type	Oil	Coal	Oil	Coal	Coal	Coal	Coal	Coal
Capacity (megawatts)	200	200	400	600	200	400	600	800
Heat rate (BTU/kwh)	8,600	10,500	8,400	10,000	10,000	9,800	9,600	9,500
Cost ($/kw)*	700	1,200	500	950	1,150	1,050	950	850
Estimated life (years)	30	34	30	34	34	34	34	34
Scheduled down time (proportion of time)	.10	.13	.10	.14	.13	.13	.14	.15
Construction lead time (years)	7	7	7	8	7	7	8	9
Fixed operating and maintenance cost ($/kw-yr)*	15.0	16.0	15.0	12.6	16.0	14.7	12.6	10.4

*In dollars of time zero.

If UTH_j is a measure of the amount and duration of capacity utilization exceeding a threshold, then user cost, in megawatts/year, is defined as:

$$(6.19) \qquad USCST_j(t) = UCMAX_j \cdot UTH_j(t)^{UCEXP_j} \cdot GCAP_j(t)$$

where UTH_j is defined in such a way that, for threshold capacity $CAPTH_j$, $0 \leq UTH_j \leq (1-CAPTH_j)/CAPTH_j \leq 1$. This function behaves as shown in Figure 6.2, where we have assumed the parameter values shown in Table 6.2.

Forced outage and variable maintenance are assumed related to one another in that it is during periods of forced outage that variable maintenance is required and takes place. Both are related to user cost in the previous time period. Thus:

$$(6.20) \quad FORC_j(t) = FOR_j + UTH_j(t-DT)^{UCEXP_j}$$

$$(6.21) \quad VMAINT_j(t) = VMNT_j \cdot USCST_j(t-DT)$$

where FOR = forced outage due to factors other than utilization rate (proportion of hours per year)

VMNT = maintenance policy (proportion of user cost)

The maintenance policy is, in principle, considered a strategy to be searched over, but for purposes of this test it is constant. The values assumed for VMNT and other parameters of these functions are shown in Table 6.2.

4. Plant Dispatching

Within each simulation period (DT = .25 year), a load duration curve is assumed and plants are dispatched in order of <u>decreasing</u> efficiency (as determined by cost of fuel needed to generate one kilowatt hour of electricity), assuming utilization does not exceed the user-specified parameter UCOPT. Any remaining load will be satisfied by further utilizing plants in <u>increasing</u> order of efficiency until the load is met or all plants are fully utilized. Then, any remaining load is met with purchased power. By setting UCOPT equal to or less than the capacity threshold, CAPTH, user cost on the more efficient plants is avoided. That is, the risk of forced outage on these base plants is minimized. If UCOPT is set to 1.0, the more efficient plants will be used to capacity before less efficient ones are dispatched.

5. Exogenous Variables

Some of the exogenous variables are assumed to be random variables reflecting the uncertainty in the problem, while others are deterministic. The deterministic random variables include the cost of capital, R, which is assumed to be a constant 11.75%/year for

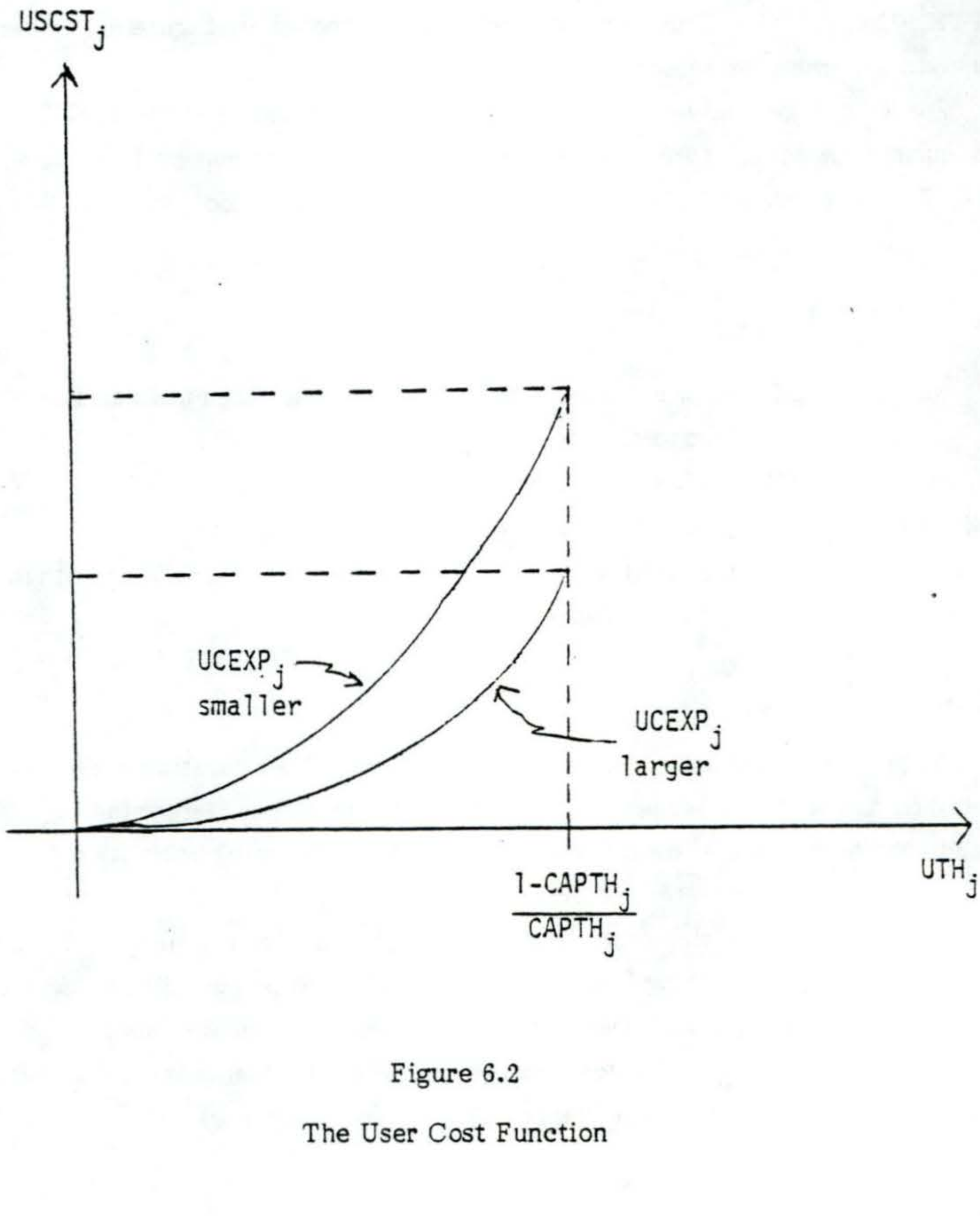


Figure 6.2

The User Cost Function

Table 6.2

Parameters of the User Cost, Forced Outage and Variable Maintenance Functions

	Initial Plants				Additional Plants			
	1	2	3	4	A	B	C	D
UCMAX (proportion/year)	1	1	1	1	1	1	1	1
UCEXP (no units)	2	2	2	2	2	2	2	2
CAPTH (proportion of capacity)	.80	.80	.80	.80	.80	.80	.80	.80
FOR (proportion of time)	.05	.10	.05	.11	.10	.10	.11	.12
VMNT (proportion of mw/year)	.99	.99	.99	.99	.99	.99	.99	.99

all plants, and the time cost or depreciation rate, D, which is assumed to be a constant 2.5%/year for all plants after the first year of acquisition and 75% during the first year. In addition, acquisition cost, ACCST, and fixed operating and maintenance cost, FOMC, are assumed to grow exponentially at 6.5 and 5.0%/year, respectively. Variable maintenance cost and disposal cost are each assumed to be 10% of acquisition cost, ACCST. Purchased power is assumed to cost 110% of the endogenous consumer price, CPR.

The random variables also are assumed to grow exponentially, with the growth rates (except in one instance) being drawn each year from given frequency distributions. The one exception is the relicensing cost, where it is the deviation from a baseline projection rather than the growth rate which is random. Data assumed for these distributions are given in Table 6.3.

6. Salvage Time and Replacement

The salvage time criterion used for coal plants [equation (6.11)] is one which assumes a plant is replaced with one identical to it (as derived in Chapter IV). It is further assumed that, when it is decided to salvage a plant, it is maintained in operation during the construction period of its replacement before actually being taken off line. Oil plants are assumed not to be replaced after retirement, except by pre-planned plants as identified by the strategy (see below).

For this version of the model, a plant's value of services, VALSER, which is key to the salvage time criterion, is not computed according to equation (6.8). Instead, as discussed above, the total gain from the operation of each plant [as given under the summation sign in the objective function shown in equation (6.6)] is allocated to the services of the durable. Further theoretical development is necessary in order to isolate the portion of the gain attributable to the plant's services from that attributable to other inputs.

A moving average of the rate of change of value of services is computed and compared to the annualized average of the value of services (or, in the case of no replacement, the control cost) to indicate the salvage time. If the moving average falls below the annualized average (or control cost), the plant is salvaged. In addition, no durable being used at more than 5 percent of its capacity is retired unless it is to be replaced.

Table 6.3

Frequency Distributions for Random Exogenous Variables

Variable	Base	Growth Rate (proportion/year) x/P(x)				
1. Oil price ($/barrel), PROIL	25	.03 10%	.09 20%	.12 40%	.15 10%	.18 20%
2. Coal price ($/ton), PRCOAL	30	.08 30%	.10 40%	.12 30%		
3. Load (mwh/yr), XLOAD	6.5 million	.01 25%	.02 30%	.03 25%	.04 10%	.05 10%
4. Deviation from baseline relicensing cost (proportion of baseline), ALPHA		-.30 10%	-.10 20%	0.0 40%	.10 20%	.30 10%

7. Control Strategies and Optimal Search

For this test case, and since control laws are assumed for utilization rates, maintenance rates, and salvage times, the control strategies subject to the optimizing search are the sizes and acquisition times of the alternative future plants under consideration. In addition, alternative control laws for utilization (i.e., dispatch rules) are tested. In total, ten strategies are tested, combining five sets of plant acquisition schedules and two dispatch rules.

Either four or five additional plants are included in each strategy, in order to add a total of 2,400 megawatts of capacity to the system over the 40-year planning horizon. Table 6.4 summarizes the combinations of plants tested.

One dispatch rule tested assumes that plants will be operated at full capacity in decreasing order of efficiency (UCOPT = 1.0). The second rule (UCOPT = CAPTH) attempts to minimize user cost by operating the more efficient plants at or below the capacity threshold unless the less efficient plants are not capable of meeting load when operating at full capacity.

For each of the 10 strategies, 20 states of nature are run, generating a cumulative distribution function of the objective function value J. The GREMP package reduces the 10 distributions to an efficient set using the criterion of stochastic dominance with respect to a function. The results are presented in the next section.

Results

This section discusses the results of the general test case with respect to (1) the selection of preferred strategies, and (2) the dynamic behavior of the simulation model.

Preferred Strategies

Each of the 10 strategies was simulated for 20 states of nature, and means and standard deviations of the objective function values were computed. Remember that, because the consumer price is endogenous as discused above, the objective is to minimize the discounted present value of costs paid by the consumer. Costs include time, user, and control costs as well as operating, maintenance, and variable input (fuel) costs. Table 6.5 summarizes the strategies and the outcome statistics.

Strategy #1 has the minimum expected (mean) cost and the second smallest standard deviation. If minimizing the expected value of the objective function were the decision criterion, Strategy #1 would be selected. The criterion used in GREMP, however--stochastic dominance with respect to a function--considers other statistical properties

Table 6.4

Plant Combinations in General Test Case Experimental Design

Combination No.	Plant No.	Capacity (mw)	Acquisition Time
1	5	400	0
	6	600	2
	7	600	12
	8	800	20
2	5	400	0
	6	400	2
	7	400	6
	8	400	10
	9	800	20
3	5	200	0
	6	200	2
	7	600	4
	8	600	12
	9	800	20
4	5	600	0
	6	600	6
	7	400	12
	8	800	20
5	5	600	0
	6	600	6
	7	600	12
	8	600	20

Table 6.5

General Test Case Results

Strategy No.	Dispatch Threshold UCOPT*	Plant Combination No.+	Statistics on Outcomes Mean++ (bill.$)	Std. Dev. (bill.$)	Rank
1	1.00	1	6.090	.220	1
2	1.00	2	6.382	.233	9
3	1.00	3	6.161	.259	2
4	1.00	4	6.167	.206	3
5	1.00	5	6.231	.242	5
6	.80	1	6.352	.418	8
7	.80	2	6.276	.243	7
8	.80	3	6.450	.420	10
9	.80	4	6.230	.299	4
10	.80	5	6.271	.295	6

* UCOPT = 1.0 means more efficient plants are used to full capacity before less efficient plants are dispatched.

UCOPT = 0.8 means more efficient plants are used at 80% capacity, in order to minimize user cost and forced outage, unless less efficient plants aren't sufficient to meet load.

\+ See Table 6.4 for definitions of the plant combinations tested.

++ Discounted present value of costs, including input, maintenance, time, user and control costs.

of the outcomes as well in relation to assumed bounds on the decision maker's absolute risk aversion function (see Chapter V), and it generally results in an efficient set of strategies rather than a unique optimum. In this general test case, the efficient set contains only one strategy--Strategy #1.

It is interesting to note, however, that the most costly strategy, Strategy #8, has an expected value of its outcome only about 6 percent greater than that of Strategy #1, yet exhibits a standard deviation almost twice as large. This suggests that, for normally distributed outcomes, the odds are about one in five (about a 19.5% probability) that Strategy #8, the "worst" in expected value, would have a better outcome than the mean of Strategy #1, a better than one in twenty chance (about a 5.7% probability) of costing less than the least cost (5.788 billion dollars) experienced in the 20 samples of Strategy #1, and an even chance of bettering the worst of Strategy #1's 20 outcomes (6.451 billion dollars). The smaller standard deviation of Strategy #1, on the other hand, gives it only about a 5 percent probability of doing worse than Strategy #8's mean, about a 78 percent chance of costing more than Strategy #8's least cost outcome (5.924 billion dollars), and virtually no chance of doing worse than Strategy #8's worst outcome (7.531 billion dollars). Pictorially, the two distributions are roughly compared in Figure 6.3, which constructs triangle distributions from the sample statistics.

It is also noteworthy that the five plant combinations tested rank differently depending on the dispatch threshold. That is, Combination #1 is best and Combination #2 is worst when UCOPT = 1.0, while UCOPT = 0.8 results in Combination #4 being best and Combination #3 being worst. On the whole, however, except for Plant Combination #2, full utilization of the most efficient plants (UCOPT = 1.0) gives lower cost results than trying to reduce user cost (and hence forced outages) for those plants (UCOPT = 0.8). As indicated above, the differences are small. Nevertheless, such differences as these are due to the dynamic impacts of the dispatch rule on utilization rates and, thence, values of services and salvage times. Furthermore, these results depend on the particular user cost functions assumed. Other, possibly more realistic, functions could have different impacts.

Model Dynamics

The dispatch rule determines plant capacity utilization by dispatching plants to full capacity from the most efficient to the least efficient until the load is met. Efficiency is calculated according to the cost of fuel, the heat content of the fuel, and the heat rate of the plant. Furthermore, plant time utilization depends on an assumed down-time for scheduled maintenance and forced outage, which is a function of user cost and maintenance regimes. Therefore, the plant utilization rates (the product of time and capacity

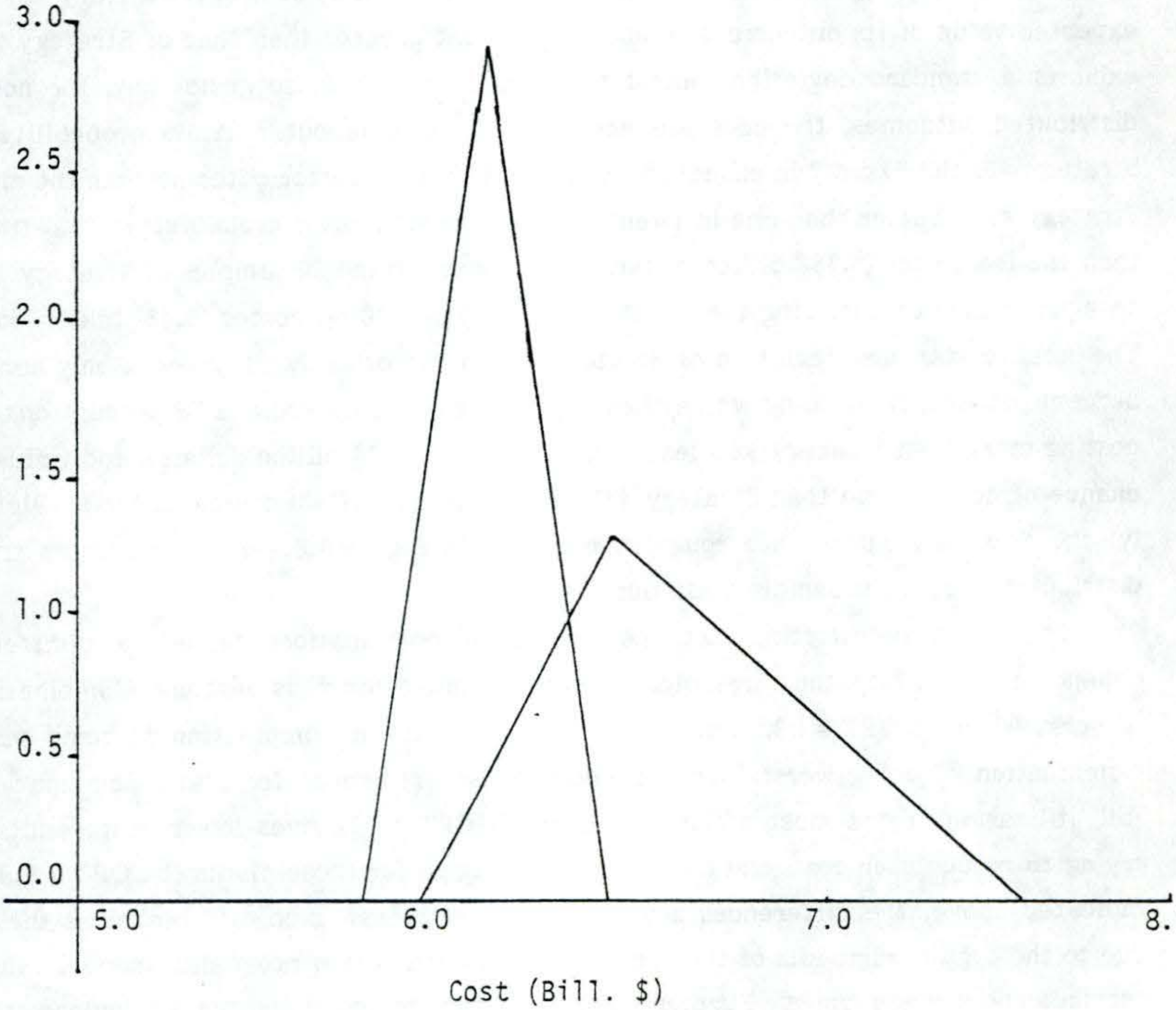


Figure 6.3

Triangle Distributions for Two Strategies utilization) provide a composite picture of the dynamic interations among the various plants of fuel costs, plant efficiencies, user costs, maintenance, and load forecasts. Figure 6.4 shows the time paths of utilization rates for the eight plants of Strategy #1 over the 40-year planning horizon.

As new plants come on line following their construction lead times (i.e., at times 7, 10, 20, and 29 for plants 5, 6, 7, and 8, respectively), existing plants experience a drop in utilization to the extent they are less efficient than the new capacity. The more efficient, new plants are then utilized to full capacity, allowing for scheduled and forced down-time, while the remaining plants steadily increase in utilization to meet the increasing load.

Plants 1 and 3 are oil plants and are assumed to be retired without replacement when the current increment to value of services, averaged over a year, does not cover control cost. Under this rule, plant 3 is retired at time 15.75. Plant 1 also meets this criteria at about the same time. However, since it is being used at a utilization rate greater than 5 percent and therefore presumably needed to meet load, it is retained. Conditions do not dictate retirement again until after plant 7 comes on line at time 20.00, and plant 1 is retired at time 24.25.

The remaining plants are coal-fired and assumed to be replaced with identical units when the current increment to value of services, averaged over a year, is less than the annualized average of value services to date, the latter representing the value of the replacement. Under this criterion, plants 2 and 4 are replaced, following lead times for construction of the replacements, at times 17.50 and 37.75, respectively.

Summary and Conclusions

We have presented a preliminary model to operationalize theoretical developments integrating the economic theory of investment and disinvestment decision making with the decision analysis theory of decision making under uncertainty. The operational model presented is oriented to decisions regarding electric power generating capacity, although it is generalizable to other durable factors of energy production and distribution.

The decision problem is conceptually cast in an optimal control framework in order to reflect the dynamic and simultaneous features of decisions, over some planning horizon, concerning: (1) when to acquire and dispose of durable productive assets (i.e., generating capacity), (2) what sizes to acquire, (3) what rate services are to be extracted from them, and (4) when and how much they are to be maintained. While it would be virtually impossible to solve such a complex, non-linear problem analytically, techniques are available for finding numerical solutions in the deterministic case. Our situation is

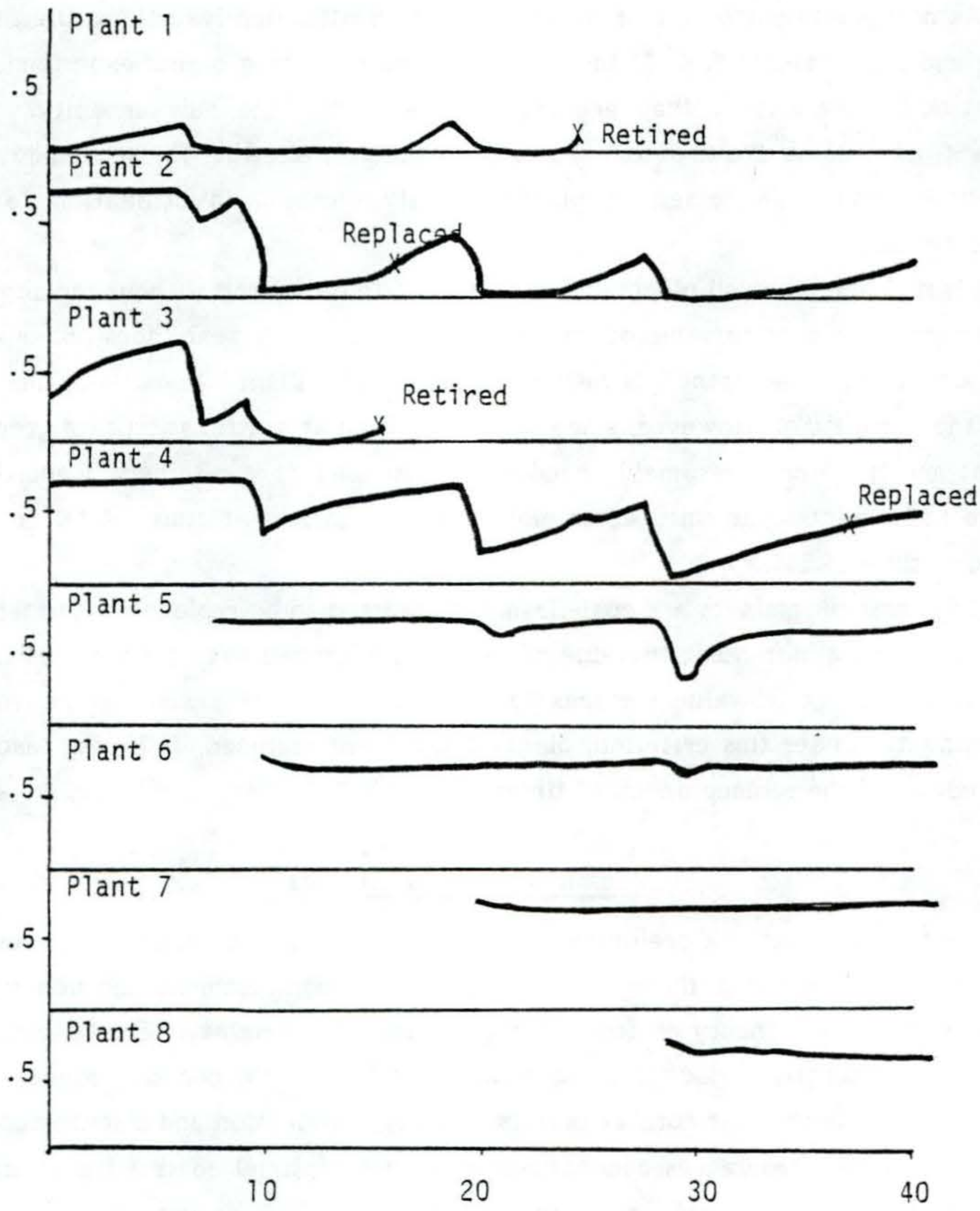


Figure 6.4

Plant Utilization Rates

decidedly not deterministic, however, because of the uncertainty part of our problem. Therefore, we have embedded the model in a software package (GREMP) that (1) systematically selects pre-specified control strategies, (2) generates a distribution of the objective function value for each such strategy through Monte Carlo simulation, and (3) identifies the efficient set of distributions (i.e., strategies), based on the criterion of stochastic dominance with respect to a function.

Finally, a general test case is constructed to test and demonstrate the range of decision features of the model. The results of this test are discussed with respect to (1) the dynamics of the problem for a single strategy and state of nature, and (2) the overall "optimal" solution, which is an efficient set of one strategy.

Experience developing the model and specifying and evaluating the test case has suggested a number of avenues for further theoretical and model development and experimentation. These are briefly enumerated below, grouped into those relating to the specifics of the test case and those concerning the basic model itself.

The General Test Case

1. Investigation of and experimentation with the parameters and functional form assumed for the stochastic dominance criterion. The one used here resulted in an efficient set of only one strategy. Was this due to the criterion function or to the peculiarities of the test problem as specified?
2. Reconsideration and, where necessary, revision of particular assumptions in the test case to increase its realism and relevance. This would include, for example:
   a. data on plant characteristics and costs;
   b. the dispatch rule (e.g., see Booth, 1971);
   c. the functional forms relating user cost, forced outage, and variable maintenance to one another and to the rate of plant utilization; and
   d. whether a control law could or should be specified relating plant acquisition time to load.

The Basic Model

3. Revision of the equation computing the value of services accrued to each plant. Currently, the total economic gain generated by a plant is attributed to the services of the plant itself. Further theoretical development is necessary for the multiple durable case, however, in order to exclude the gain attributable to other inputs used in conjunction with the plant's services.

4. Reconsideration of the criteria determining the optimal salvage time for a plant, i.e., STIME in equation (6.11).
   a. The derivation of the criterion for replacement with an identical unit assumes the value of services is well-behaved over time; specifically, that the second-order conditions for maximization are met. This is not the case in general, primarily due to the interdependency of plants through the dispatch rule and other constraints imposed on a plant's operation. That is, the criterion was developed for a single durable. Further theoretical development is necessary to derive a criterion which considers the interactions among the multiple durables of a single firm which is constrained to produce a given level of output over time.
   b. The criteria implemented here were derived assuming either no replacement or that the durable would be replaced with an identical unit which faces identical price, cost, and use patterns over time. Alternative criteria need to be derived for other cases, such as replacement with altogether different types of units.
5. Implementation of an algorithm to find a numerical solution to the deterministic optimal control problem. A software package to do so could be appended to the GREMP system. Such a solution would be instructive and useful to check the logic and realism of the problem specification and to compare with the set of efficient strategies resulting from the GREMP algorithm.

CHAPTER VII

A COMPARATIVE TEST OF THE EMPIRICAL MODEL

Michael H. Abkin and Gary R. Ingvaldson

Introduction

One way to evaluate a new approach or model, particularly one intended to be useful in practical, decision-making settings, is to compare its content and performance with those of an existing model currently used for similar purposes. This chapter reports on the results of such a test, wherein the empirical model presented in Chapter VI was compared with the minimum revenue requirements method* in the context of an actual decision faced by the Cosumers Power Company of Jackson, Michigan.+

In 1969, Consumers Power Company decommissioned and divested itself of two hydroelectric stations, the Boardman and Sabin dams, on the Boardman River in Grand Traverse County, Michigan. The energy economics of that time did not justify the labor expense of operating and maintaining those plants, which produced only a tiny part (2.2 mw and 0.9 mw, respectively) of Consumers' power output. Then, in 1977, under radically different and changing relative prices for energy and labor, Grand Traverse County and Traverse City asked Consumers Power to consider reacquiring the dam sites and reactivating the hydro stations to generate power which the City and County could purchase to help meet their projected load growth. Consumers Power studied the proposal and decided against pursuing it.

In order to make a real-world application of the theoretical and practical investment/disinvestment models presented in Chapters IV and VI, Consumers Power Company supplied us with the background data used in the above study. The objective was to have a problem simple enough for this first, experimental test of the model and for which the model's application and results could be compared with those of an existing model, namely, the revenue requirements method commonly used in the electric power industry.

An additional, invaluable source of data and information for this test was the final report (Joint Venture, 1979)++ of a feasibility assessment for renovating five dams on the Boardman River, of which the Boardman and Sabin are two.

---

*See Chapter II for a description and evaluation of the minimum revenue requirements method. Also, see Boris (1978).

+We are indebted to Jim Parker, Angelo Muzzin, Ron Calcaterra, and others at Consumers Power Company for their active interest, cooperation, and patience in the course of this study.

++We are grateful to Bill Strom of Traverse City Light and Power for making this study available to us.

The next section of this chapter summarizes the data and other assumptions used in this test application as derived from the above two sources. The third section, then, discusses the results of the analysis, and, finally, concluding remarks are presented.

Assumptions

The presentation of assumptions here parallels that given in Chapter VI for the general test case of the investment/disinvestment/utilization/maintenance decision model. Variable names are also those used in Chapter VI.

1. Planning Horizon and Discount Rate

$T = 50$ years; $\rho = 11.75\%$

2. Plants To Be Considered

Only two plants are defined in the analysis: the Boardman and Sabin hydroelectric plants, neither of which is assumed to exist initially. Therefore, $k = 0$ and $n = 2$. At 2.2 mw and 0.9 mw, respectively, these plants would represent an increase of only about one-twentieth of one percent of Consumers Power's 6000 mw capacity. Incorporating the existing Consumers Power system into a simultaneous analysis of these two sites would be impractical, and any differences in the objective function arising from various strategies with respect to them would be negligible. Therefore, the analysis implicitly treats the existing system as a zero reference point. Further, in keeping with this approach, no load forecast or load constraint is assumed, and the energy output of the plants is priced at the replacement power cost.

3. User Cost, Forced Outage, and Variable Maintenance

The same functional forms and relationships are assumed here for these variables as in the general test case [equations (6.19)–(6.21)]. However, given the nature of hydroelectric power generation, wherein less regular maintenance is required and intensive utilization requires less variable maintenance than in steam plants, the parameter values for these functions are chosen to reduce their magnitude. Specifically, only two percent down-time each is assumed for both scheduled maintenance, SCHED, and basic forced outage, FOR, i.e., forced outage due to factors other than the utilization rate. Furthermore, the capacity utilization threshold below which no user cost is incurred, CAPTH, is assumed to be a high 90 percent. Thus, the maximum capacity loss rate, i.e., if the plants are run at full capacity for the maximum available time, will be only

$(0.1/0.9)^2$ = 1.25 percent per year. It is unlikely to actually approach anything like this, however, because of the dispatch rule employed, as discussed below.

4. Plant Dispatching

Ideally, with zero fuel (water) cost, one would run these plants all the time as base plants. However, with virtually no pond storage capacity behind either the Boardman or Sabin dams, the power output is essentially limited to the run of the river. Figure 7.1 is a flow duration curve for the Boardman River, as derived in Joint Venture (1979). The vertical axis is in cfs (cubic feet per second) per square mile. This curve is used in the present analysis to determine water flow, WFLOW, in cfs by assuming the mean of 1.064 cfs/sq. mi. corresponds to 285 cfs, given elsewhere in that study for the Boardman and Sabin dams, and the maximum flow of 3 cfs/sq. mi. corresponds to 683 cfs. Furthermore, as also specified in that study, it is assumed the plants cannot operate below 180 cfs. Given the water flow rate, then, the power output in kilowatts is computed by PROD = WFLOW · HEAD/14.75, where the Boardman and Sabin dams have heads of 41 and 21 feet, respectively.

5. Exogenous Variables

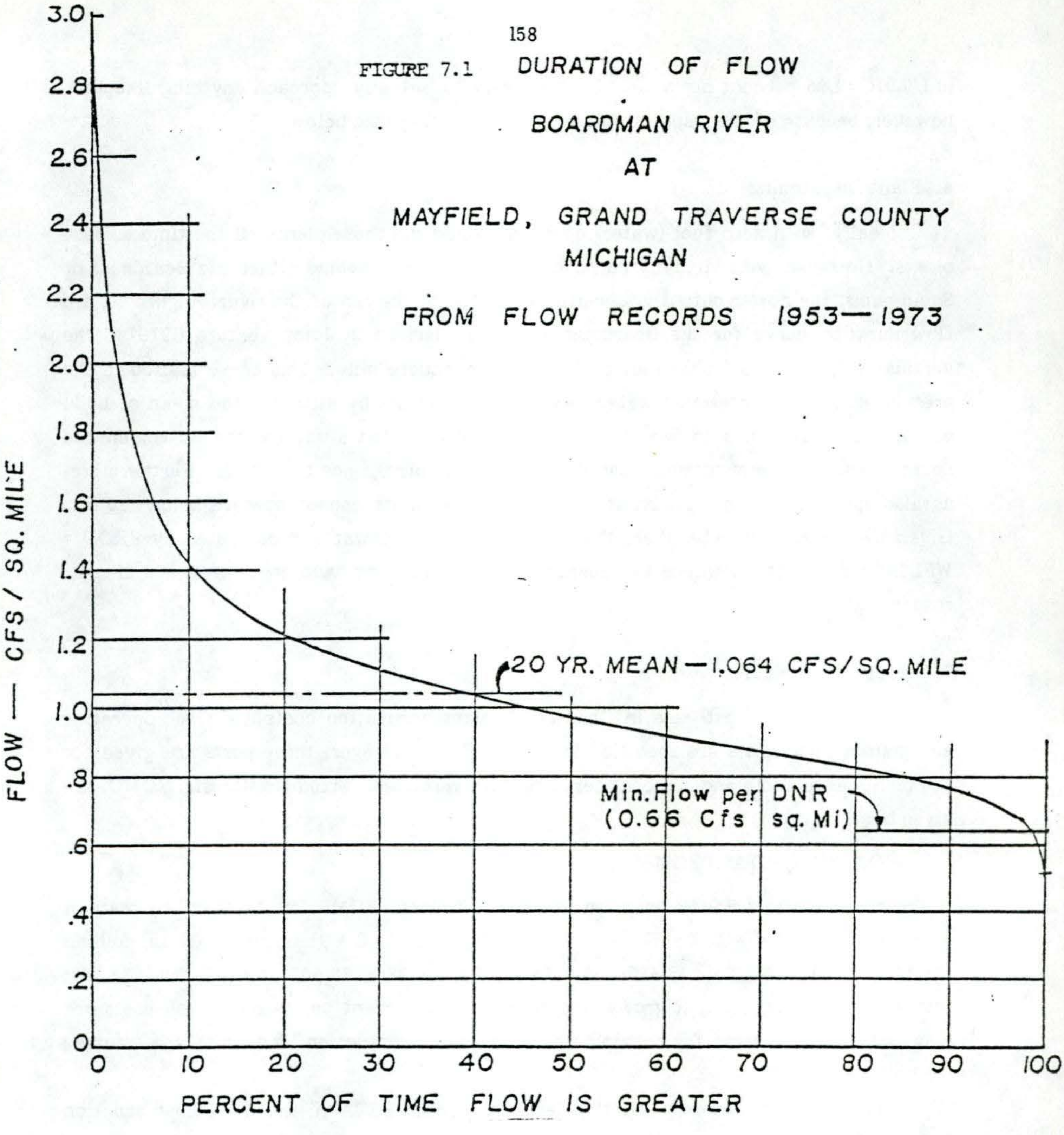
In the general test case in Chapter VI, plant acquisition costs and fixed operating and maintenance costs are specified in $/mw. Here, however, these costs are given for the plants as a whole and are considered random variables. Fixed O&M costs, FOMC, are given by:

FOMC(t) = BETA(t) FOMB(t)

where the basic cost FOMB, is given in current dollars. FOMB for the first ten years of operation (from t = ACQT + TLEAD to t = ACQT + TLEAD + 9) is given as, in dollars: 128,244; 59,493; 44,705; 55,655; 58,477; 61,300; 66,380; 78,007; 69,766; and 72,363; respectively. After that, it grows at a rate of 6.25 percent per year. These costs are assumed to be the same for both plants. Two-year construction lead times are assumed for both plants.

Plant acquisition costs and the deviation factor BETA in the O&M cost equation above follow the cumulative distribution functions shown in Table 7.1. Similarly, replacement power cost follows the frequency distribution given in Table 7.2. After 1991, the whole distribution is assumed to increase at a 6 percent annual rate.

FIGURE 7.1



Source: Figure taken from Joint Venture (1979).

Table 7.1

Cumulative Distribution Functions for
Plant Acquisition Costs and O&M Costs

XRAN = RANF(-1)	0.00	0.25	0.50	0.75	1.00
ACCST (Boardman) (mill. $)	1.43	1.47	1.50	1.53	1.65
ACCST (Sabin) (mill. $)	1.33	1.37	1.40	1.43	1.54
BETA (proportion)	0.95	0.98	1.00	1.02	1.10

Table 7.2

Frequency Distribution of Replacement Power Cost
($/mwh)

Year	P=.1	P=.2	P =.4	P =.2	P =.1
1980	35.1	38.9	45.9	52.2	58.6
1981	34.9	40.4	45.8	55.6	81.8
1982	40.7	49.4	58.0	77.7	95.6
1983	31.0	40.1	51.0	67.3	87.3
1984	31.7	42.5	57.4	76.6	105.5
1985	36.0	51.1	73.8	99.3	146.7
1986	39.5	56.3	80.1	108.8	159.3
1987	42.0	55.8	78.8	109.2	158.0
1988	46.5	67.2	93.1	130.4	191.5
1989	22.9	31.7	44.3	62.0	99.7
1990	54.7	74.0	103.4	146.0	254.4
1991	67.1	87.2	120.8	176.8	326.8

6. Control Strategy

With plant utilization and maintenance determined by the decision rules discussed above, and in order to compare the methodology and results with the Consumers Power study, only one strategy is evaluated for 20 states of nature. That strategy is to acquire both dam sites--2.2 m for the Boardman and 0.9 mw for the Sabin--at time zero (1978). The results are presented in the next section.

Results and Comparative Discussion

Since the plants' production is valued at the replacement power cost, which was projected independently by Consumers Power Company to represent the marginal cost of power, the objective in this analysis is to maximize the discounted present value of economic gain [equation (6.6)]. This is in contrast to the cost minimization objective of the general test case in Chapter VI.

For the unique strategy tested, the expected value of the economic gain is $709,000 over the 50-year planning horizon, with high and low values of $1,088,000 and $476,000, respectively, and a large standard devitation of $141,000. The mean is rather small relative to the $3.1 million initial investment, representing only about a 2.7 percent levelized annual return over the 50-year horizon. The standard deviation, at 20 percent of the mean, is large, indicating a risky venture. The riskiness is further augmented by the long payback period. For one of the states of nature, the one giving the third highest gain ($814,000), the objective function did not become positive until year 28.25. A more average (i.e., lower gain) result presumably would have an even later break-even point.

Consumers Power's original analysis, with the minimum revenue requirements method and using the same data, had similar results--a marginally positive gain with a long payback period. The implied riskiness plus other considerations not included in this analysis (such as the costly legal requirement to install fish ladders) dictated the decision not to reactivate the dams.

Computationally, the two approaches are very similar to one another. In the Boardman study, both looked at the discounted present value of economic gain and came to similar conclusions. In the general test case of Chapter VI, the investment/disinvestment/utilization/maintenance model was led to consider a cost minimization objective. This is justified because consumer prices are regulated in such a way as to automatically pass costs through to the customer, including an allowed rate of return on capital which compensates for the opportunity cost of capital. This is identical to the minimum revenue requirement's objective of selecting among alternative projects so as to minimize the revenues from customers required to pay for the initial and recurring costs of the project (see Chapter II).

With respect to dealing with uncertainty, again there is little or no difference. The minimum revenue requirements method can and often does use Monte Carlo analysis to capture uncertainties in cost and other projections. Moreover, the GREMP approach of finding a risk-efficient set of decision options by using the criterion of stochastic dominance with respect to a function can certainly be used with a minimum revenue requirements model if desired.

The major differences between the two approaches are in the definition and computation of costs and revenues and the determination of salvage time. In the minimum revenue requirements method, a useful life is assumed and a "fixed charge rate" computed accordingly, encompassing a minimum rate of return, a standard book depreciation annuity, and allowances for insurance and federal, state, and local taxes (Boris, 1978). The fixed charge rate is based on the initial cost of the project (or durable asset). To this are added projections of future operating expenses. Typically, revenues are ignored.

In the model presented here, on the other hand, both costs and revenues are considered over the life of the durable, and that life itself is not assumed but rather determined endogenously as an economic decision. Revenues are necessary not only for the economic gain objective function (except for regulated utilities, as discussed in Chapter VI) but also to compute the value of services attributable to the durable, an important criterion in deciding whether to acquire the durable and when to disinvest in (or salvage) it. On the cost side, rather than considering the cost of capital and depreciation in a levelized fixed charge rate, these are explicitly accounted over the life of the durable by separately identifying time costs, user costs, and control costs, as defined in Chapter IV. Furthermore, these costs are based on the replacement rather than the initial cost of the durable.*

Conclusion

Despite the differences cited above between minimum revenue requirements and the approach presented here, application of the two models to the Boardman analysis yielded similar results. This will not necessarily be the case in general, however.

As we developed our model and its Boardman application, and as our Consumers Power Company advisors became more familiar with the model through critical evaluation, we came to realize that the Boardman study, while providing a simple case history which had been completed and for which the data were readily available, could not fully

*We recognize, however, that tax and regulatory policies generally dictate the use of initial cost in computing depreciation and rates of return.

test the comprehensive decision features (investment, disinvestment, utilization, and maintenance) of the model. Essentially, the Boardman represents a small hydroelectric power plant with no pond storage and generating less than one percent of the system's power output. Its utilization rate is determined solely by the run of the river, and it requires little maintenance.

The conclusion is that the proposed model, while suitable for partial or project analysis, has the capability for applications to strategic capacity planning on a system-wide basis. A test case at this level, perhaps as a component of a corporate planning model (e.g., Selby, 1979), might better explore the limits of the model. The minimum revenue requirements method, however, is not well-suited to this strategic level of analysis. Therefore, it is suggested that an alternative approach, such as the over/under model (EPRI, 1979), be identified for further comparison purposes.

CHAPTER VIII

RECOMMENDATIONS FOR FUTURE WORK

Lindon J. Robison and Michael H. Abkin

Introduction

In the process of building a theory, researchers must often begin with analytic models that are simplifications of the world they model. They ignore, in the process, many elements of the real world that may be important. Such has been the case in this study--many important elements in the theory of investment/disinvestment were not modeled. So, in any future work on investment/disinvestment theory and the analysis of potential supply responses under uncertainty of a regulated industry, it is important to state (1) what was done; (2) what was not done; and (3) what needs to be done. The executive summary at the beginning of this report summarized what was done in this study. We now discuss the areas this study did not analyze and where additional research is needed.

"Conglomerate" Durable Analysis

The weakness of our model is that investment/disinvestment problems are seldom of the single-durable type analyzed in this report. Instead, we find "conglomerate" durables, each producing interrelated and interdependent services. The durable used throughout Chapter IV as an example--a passenger car--illustrates this point. The service delivered by the car is transportation, but its ability to deliver this service depends on services provided by its component parts, including the battery, the alternator, the radiator, the tires, the starter motor, the lubricants, etc., each of which is itself a durable. The durable we defined as a car is, in fact, a collection of durables--a conglomerate. Moreover, the life of the car cannot be determined without recognizing this interdependent nature of services obtained from all durables. At some point, the failure, or imminent failure, of the durables included in the conglomerate durable may give rise to the decision to disinvest in the car. What this paper has not done is to provide the analytic framework for making such a decision. This is an area that demands attention.

Durables in Parallel and Durables in Series

The investment/disinvestment analysis of conglomerate durables is closely related to still another class of investment/disinvestment problem--durables in parallel or in series. "Durables in parallel" is the investment/disinvestment problem that occurs when more than one durable provides the same service. Base and peaking units are examples. They are durables that may simultaneously produce the same service, but, because of differences between the two in time and user costs, they are not perfect substitutes in the production process.

This report did not build a theory to analyze durables in parallel. Introducing a second durable in parallel with the first adds an opportunity cost of producing from the first durable--the returns given up by not using the idle durable in production. The model presented in this study, which approximated this cost by control costs, needs to be improved to analyze such a result. Obtaining such a theory would allow us to perform still another task--deriving optimal dispatch rules for using durables in parallel.

"Durables in series" poses another challenging problem. When the output of one durable provides services to still another, how are their respective contributions measured? Or, how is the investment/disinvestment in one related to the other? A familiar example in agriculture is the decision farmers make to replace a tractor, realizing that, by so doing, the services from an entire line of machinery complements may be affected. Electric utilities experience this problem in the series of generating plants, transformers, and transmission lines.

Baquet did analyze such a problem and proposed an iterative approach for resolving it. However, it is not clear that the measurement of contributions from the durable was appropriate. Moreover, in place of replacement opportunity costs, he used a very long planning horizon. Still, any extensions in this work could benefit from his pioneering study. The result to be hoped for would be an analytic framework that could be implemented with computer programs that would permit analysis of much more complicated environments than the one of this study.

Regulations, Taxes, and Inflation

As the disciplinary model developed in this paper is adapted to include more relevant features of the decision environment--i.e., becomes more subject-matter oriented--three areas come to mind: regulations, taxes, and inflation. All three are relevant for public utilities. Regulations are a fact of life. Rates of return are tied to the industry's rate base valued at book value, but which, because of inflation, is no longer an appropriate measure. This issue has only been touched on in this report (Chapter IV). Linking rates of return to the utility's rate base, valued in terms of book rather than market value, may lead to cash flow problems for the firm. On the other hand, accounting for capital gains, a non-cash return, becomes a related problem.

Many articles have been written about how linking returns to the firm's capital base leads to over-capitalization (Averch and Johnson, 1962). Although several variants on this theme have been proposed, one that deserves attention is an analysis that links inversely the probability that a capital investment will be included in the utility's rate base with the difference to the customer of operating with the new investment versus some least cost alternative. The result may very well be a model that produces least-cost-to-the-customer capitalization rates.

Uncertainty

In every phase of the analysis of durable investments, disinvestments, and use is the presence of uncertainty. As if the uncertainty of demand associated with weather and other factors, the uncertainty associated with the availability and costs of inputs, and the uncertainty of changing regulations--as if these were not sufficient cause for decision-making headaches, utilities face still another potential source of uncertainty, namely, intermittent supply sources.

Supporting the adoption or at least the exploration of alternative energy sources is the Public Utilities Regulation and Policy Act (PURPA), one provision of which requires utilities to buy back excess energy from customers. One potential source of energy that customers could use, as well as sell back to utilities, is excess energy produced from wind-powered generators. This would create an uncertain supply source for utilities. The ramifications of PURPA on an industry required to purchase energy from an intermittent and uncertain source, as well as meet an uncertain time varying demand, needs examination. Such a study should also research the impact of energy-conserving technologies on utilities' capacity requirements needed to meet uncertain demands, given reliability constraints.

Extend Existing Economic Theory

Analytic models developed in economics can be variously classified by the market in which prices are determined and by the presence or absence of uncertainty. In almost all analytic results, however, the theories for non-durable inputs are divisible in use and in acquisition. This is convenient because it allows the models to ignore the problem of disinvestment (the asset is purchased in exactly the quantity that will be used), and it simplifies the acquisition question to the standard marginal rule: acquire input until the value of its services equals its cost.

These economic models, while helpful, ignore a much broader class of problems: input investment decisions of non-durable inputs not available for acquisition in infinitely divisible quantities and, in some cases, whose services are not available in divisible quantities. For those durables which provide divisible services, the question of what level of services to extract can be answered using marginal analysis tools--but only if intertemporal dependencies are recognized. Then, the larger question--should the durable be acquired, and, if so, when should it be replaced--must still be answered.

This study has provided only the introduction of a framework for such an analysis. It establishes a link between divisible disciplinary models of economics and lumpy investment/disinvestment models of the financial analysts. Still, there remains a wealth of models and results to be examined. Such everyday problems as how changing prices of variable inputs used by a durable affect the durable's value to the firm are questions of interest which could provide major advancements in our understanding of durable investment/disinvestment decisions.

The Simulation Model

This study has examined the simplest of all durable investment/disinvestment problems: the replacement of a single durable. Yet, even the simplest investment/disinvestment model led to what most would agree were complicated analytics, particularly with the introduction of uncertainty. Still, the effort was helpful in that a standard format for analyzing investment/disinvestment problems was identified (Chapter IV). For implementation in practical decision making, this format was cast into an optimal control framework in conceptualizing the structure of a simulation model (Chapter VI). As simplifying assumptions--which were made to render tractable the analytical models of the theory--are relaxed to make the analysis more relevant in practice, simulation models and computers become essential additions to the analyst's tool kit.

Therefore, developments of the simulation model are needed to parallel the theoretical developments recommended above. The resulting model, then, would be of increasing usefulness to utility planners as it incorporates more realistic features of the world they face. At the same time, the model would remain of intermediate complexity with respect to the manpower and computer resources required to use it. That is, it would not replace the system or corporate planning models used by utilities but instead would provide a lesser cost means of evaluating and screening options during the initial stages of planning.

Finally, although the simulation model has been conceptualized in an optimal control framework, no optimal solution is actually found. Rather, the results of assuming alternative decision rules and decision actions are compared and, in the stochastic, uncertainty case, an efficient set of options is identified. From a utility theory point of view, as well as a practical one, this is preferable to a single "optimal" solution (Chapter V). Nevertheless, such a solution would be instructive and useful--in the deterministic, or even the stochastic, case--to check the logic and realism of both the problem specification and the efficient set of strategies. Therefore, a further worthwhile development would be implementation of an algorithm, as part of the software package, to find a numerical solution to the optimal control problem.

REFERENCES

Abkin, M. H. and G. R. Ingvaldson. "A Software Package for Analysis of Investment and Disinvestment Under Uncertainty." Technical Documentation and Programmer's Guide, Staff Paper No. 80-72, Department of Agricultural Economics, Michigan State University, 1980.

Anderson, J. R. "Programming for Efficient Farm Planning Against Non-Normal Risk." Australian Journal of Agricultural Economics 19(1975):94-107.

Arrow, K. J. Aspects of the Theory of Risk Bearing. Helsinki, 1964.

_________. Essays in the Theory of Risk Bearing. Chicago: Markham Publishing Co., 1971.

Averch, H. and L. L. Johnson. "Behavior of the Firm Under Regulatory Constraint." American Economic Review 52(1962):1053-1069.

Baquet, A. E. "New Investment and Disinvestment Theory" in Capital Investment Decisions Workshop Proceedings, EPRI WS 77-18, 1977.

_________. "The Theory of Economic Thought on the Theory of Investment and Disinvestment" in Capital Investment Decisions Workshop Proceedings, EPRI WS 77-18, 1977a.

_________. "A Theory of Investment and Disinvestment Including Optimal Lives, Maintenance and Usage Rates for Durables." Unpublished Ph.D. dissertation, Michigan State University, 1978.

Behrman, J. Supply Response in Underdeveloped Agriculture. North Holland Publishing Co., 1968.

The Bell Journal of Economics 7(1976):197-242.

Booth, R. R. "Power Systems Simulation Model Based on Probability Analysis." IEEE PICA Conference, May 1971.

Borch, K. "A Note on Uncertainty and Indifference Curves." Review of Economic Studies 36(1969):1-4.

Boris, W. R. "Practical Industry Perspectives on Investment Decisions Affecting Energy Supplies--The Electric Utility Industry" in Capital Investment Decisions Workshop Proceedings, EPRI WS 77-18, May 1978.

Conklin, F. S.; A. E. Baquet; and A. N. Halter. "A Bayesian Simulation Approach to Estimating Value of Information: An Application to Frost Forecasting." Technical Bulletin 136, Agricultural Experiment Station, Oregon State University, 1977.

Edwards, C. "Resource Fixity, Credit Availability and Agricultural Organization." Unpublished Ph.D. dissertation, Michigan State University, 1958.

_________. "Resource Fixity and Farm Organization." Journal of Farm Economics 41(1959):747-759.

Electric Power Research Institute. "Over/Under Capacity Planing Model." Final Report of EPRI Project No. 1107 EA/1117, July 1979.

Energy, Department of. The Integrating Model of the Project Independence Evaluation System. June 1, 1977.

Feldstein, M. S. "Mean Variance Analysis in the Theory of Liquidity Preference and Portfolio Selection." Review of Economic Studies 39(1976):5-12.

Friedman, M. and L. J. Savage. "Utility Analysis of Choices Involving Risk." Journal of Political Economy 56(1948):279-304.

Georgescu-Roegen, N. "The Economics of Production." American Economic Review 61(1971):1-9.

__________. The Entropy Law and the Economic Process. Cambridge: Harvard University Press, 1971a.

Gilbert, R. J.; D. M. G. Newberry; and J. E. Stiglitz. "An Overview of the Economic Theory of Uncertainty and Its Implications for Energy Supply." EPRI Special Report EA-58-6-SR, January 1978.

Gilbert, R. J. and J. E. Stiglitz. Effects of Risk on Prices and Quantities of Energy Supplies. EA-700, Vol. 2, final report prepared for Electric Power Research Institute, May 1978.

Hadar, J. and W. R. Russell. "Rules for Ordering Uncertain Prospects." American Economic Review 59(1969):25-34.

Hanoch, G. and H. Levy. "The Efficiency Analysis of Choices Involving Risk." Review of Economic Studies 36(1969):335-346.

Hartman, R. "Factor Demand with Output Price Uncertainty." American Economic Review 66(1976):675-681

Holthausen, H. "Input Choices and Uncertain Demand." American Economic Review 66(1976):94-103.

Idachaba, F. "Rate of Use, Investment and Disinvestment." Unpublished manuscript, Department of Agricultural Economics, Michigan State University, 1972.

Jeynes, P. H. Probability and Economic Choice. Ames, Iowa: Iowa State University Press, 1968.

__________. "The Minimum Revenue Requirements Discipline (MRRD)." Proceedings 1978 Fall Industrial Engineering Conference, pp. 113-122.

Johnson, G. L. "The State of Agricultural Supply Analysis." Journal of Farm Economics 42(1960):435-452.

Johnson, G. L. and L. S. Hardin. Economics of Forage Valuation. Station Bulletin 623, Agricultural Experiment Station, Purdue University, April 1955.

Johnson, G. L. and C. L. Quance, eds. The Overproduction Trap in U.S. Agriculture. Baltimore: The Johns Hopkins University Press, 1972.

Joint Venture of City of Traverse City, Michigan, and Grand Traverse County, Michigan; and Emery and Porter, Inc. (consultant). "Feasibility Study for Boardman River Hydroelectric Power." Final Report to U.S. Department of Energy under Cooperative Agreement No. EW-78-F-07-1792, February 22, 1979.

Joskow, P. L. "Contributions to the Theory of Marginal Cost Pricing." Bell Journal of Economics 7(1976).

Just, R. E. "An Investigation of the Importance of Risk in Farmers' Decisions." American Journal of Agricultural Economics 56(1974):14-25.

Keady, F. D. "Investment Policy in the Coal Industry" in Capital Investment Decisions Workshop Proceedings, EPRI WS 77-18, 1977.

Keynes, J. M. The General Theory of Employment, Interest and Money. New York: Harcourt, Brace, 1936.

Kihlstrom, R. E. and L. J. Mirman. "Risk Aversion with Many Commodities." Journal of Economic Theory 8(1974):361-388.

King, R. P. "Operational Techniques for Applied Decision Analysis Under Uncertainty." Unpublished Ph.D. dissertation, Department of Agricultural Economics, Michigan State University, 1979.

Kirk, D. E. Optimal Control Theory: An Introduction. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970.

Knight, F. H. Risk, Uncertainty and Profit. New York: Houghton Mifflin Co., 1921.

Knowles, G. J. "Estimating Utility Functions." Paper presented to Western Regional Committee on Market Risk in Agriculture, Tucson, Arizona, January 16-18, 1980.

Levy, H. and J. Paroush. "Toward Multivariate Efficiency Criteria." Journal of Economic Theory 7(1974):129-142.

Lewis, W. A. Overhead Costs. London: Allen and Unwin, 1949.

Lin, W. "Measuring Aggregate Supply Response Under Instability." Paper presented at AAEA annual meetings, San Diego, California, August 1, 1977.

Lin, W. W. and H. S. Chang. "Specification of Bernoullian Utility Function in Decision Analysis." Agricultural Economic Research 30(1978):30-36.

Luenberger, D. G. Introduction to Linear and Nonlinear Programming. Reading, Massachusetts: Addison-Wesley Publishing Co., 1973.

Markowitz, H. Portfolio Selection: Efficient Diversification of Investments. New York: Wiley and Sons, 1959.

Marshall, A. Principles of Economics. New York: The Macmillan Co., 1948 (first published in 1890).

Meyer, J. "Choice Among Distributions." Journal of Economic Theory 14(1977):326-336.

__________. "Second Degree Stochastic Dominance with Respect to a Function." International Economic Review 18(1977a):477-487.

Microeconomic Associates. "An Overview of the Economic Theory of Uncertainty and Its Implications for Energy Supply," EPRI EA-586-SR, January 1978.

Neal, A. C. "Marginal Cost and Dynamic Equilibrium." Journal of Political Economy 50(1942):45-64.

Officer, R. R. and A. N. Halter. "Utility Analysis in a Practical Setting." American Journal of Agricultural Economics 50(1968):257-277.

Perrin, R. K. "Asset Replacement Principles." American Journal of Agricultural Economics 54(1972):60-67.

Pratt, J. W. "Risk Aversion in the Small and in the Large." Econometrica 32(1964):122-136.

Robison, L. J. and P. J. Barry. "Portfolio Adjustments: An Application to Rural Banking." American Journal of Agricultural Economics 59(1977):311-320.

Robison, L. J. and J. R. Brake. "Application of Portfolio Theory to Farmer and Lender Behavior." American Journal of Agricultural Economics (forthcoming).

Robison, L. J. and G. Carmen. "Aggregate Risk Models and Market Equilibrium." Paper presented at the Western Regional Research Committee, San Francisco, California, January 1979.

Robison, L. J. and R. P. King. "Specification of Micro Risk Models for Farm Management and Policy Research." Agricultural Economics Report 349, Department of Agricultural Economics, Michigan State University, December 1978.

Rothschild, M. and J. E. Stiglitz. "Increasing Risk II: Its Economic Consequences." Journal of Economic Theory 3(1971):66-84.

Ryan, T. J. "Supply Response to Risk: The Case of U.S. Pinto Beans." Western Journal of Agricultural Economics 2(1977):35-43.

Sandmo, A. "On the Theory of the Firm Under Price Uncertainty." American Economic Review 61(1971):65-73.

Selby, J. D., et al. "New Dimensions in Utility Planning." IGT Symposium on Energy Modeling, Colorado Springs, Colorado, August 1979.

Smith, G. W. Engineering Economy: Analysis of Capital Expenditures. Ames, Iowa: Iowa State University Press, 1968.

Smith, K. "On Risk and Optimal Rates of Resource Utilization in a Theory of the Firm." Unpublished Ph.D. dissertation, Northwestern University, 1968.

__________. "The Effect of Uncertainty of Monopoly Price, Capital Stock and Utilization of Capital." Journal of Economic Theory 1(1969):48-59.

Stobaugh, R. and D. Yergin, eds. Energy Future: Report of the Energy Project at the Harvard Business School. New York: Random House, 1979.

Traill, B. "Risk Variables in Econometric Supply Response Models." American Journal of Agricultural Economics 60(1978).

Tsiang, S. C. "The Rationale of the Mean-Standard Deviation Analysis, Skewness, and the Demand for Money." American Economic Review 62(1972):354-372.

Turnovsky, S. J. "Production Flexibility, Price Uncertainty, and Behavior of the Competitive Firm." International Economic Review 14(1973).

Vanston, L. K. and M. L. Baughman. "The Representation of the Electrical Utility Sector in the DOE/MREFS: Overview and Comment." Center for Energy Studies, University of Texas at Austin, February 1979.

Wenders, J. T. "Peak Load Pricing in the Electric Utility Industry." Bell Journal of Economics 7(1976).

Whitmore, B. A. "Third Degree Stochastic Dominance." American Economic Review 60 (1970):457-459.

Winfrey, R. Statistical Analysis of Industrial Property Retirements. Ames, Iowa: Iowa State University Press, 1935.

Zadeh, L. A. "Outline of a New Approach to the Analysis of Complex Decision Processes." IEEE Transactions on Systems, Man, and Cybernetics, SMC-3 (1973): 28-44.