

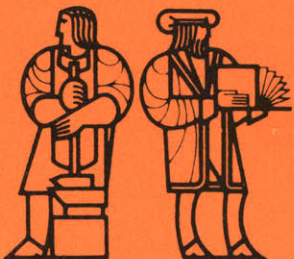
ENERGY LABORATORY

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

SPOT PRICING
OF PUBLIC UTILITY SERVICES

by
Roger E. Bohn

MIT Energy Laboratory Report MIT-EL 82-031
May 1982



Spot Pricing of Public Utility Services

by

Roger Eric Bohn

Submitted to the MIT Sloan School of Management on May 28, 1982
in partial fulfillment of the requirements for the degree
of Doctor of Philosophy

ABSTRACT

This thesis analyzes how public utility prices should be changed over time and space. Earlier static and non spatial models of public utility pricing emerge as special cases of the theory developed here. Electricity is emphasized although the models can be used for natural gas and other public utilities.

If the transactions costs of price changes were zero, optimal prices should be changed continuously as supply and demand conditions change. Such prices are referred to as "full spot" prices. The full spot price of electricity at any point and time depends on total demand, availability of generating units, short run marginal operating costs of generators, the spatial locations of all supplies and demands, and the configuration of the transmission and distribution system. Since all of these are stochastic, so are full spot prices. Optimal "wheeling charges", i.e. price differences between points, also vary stochastically.

In practise actual prices must be changed discreetly, and are therefore only approximations to full spot prices. Price changes are of two basic types. Predetermined price changes are adequate to respond to anticipated changes in conditions. Price recalculations are needed to respond to unanticipated changes. The optimal timing and mix of recalculations and predetermined changes depend on: the transactions costs of each type of change; the stochastic and deterministic rates of change of full spot prices; and the ability of customers and suppliers to change behavior in response to different price patterns. Conventional time-of-use rates recalculate prices systematically only at occasional rate hearings, and change them only a few times a day. Such prices deviate greatly from full spot prices, for many utilities.

The thesis models customer behavior under full spot and other time varying prices, and discusses the types of customers likely to get the largest benefits from full spot pricing. The final chapter simulates behavior by four customers under six rates, from flat prices

calculated a year in advance, to full spot pricing. The gross social welfare benefits of full spot pricing, before transactions costs, are three to ten times the benefits of conventional time-of-use pricing. The gross benefits of full spot pricing are less than ten percent of the customers' total energy costs. For small customers this may not be enough to counterbalance higher transactions costs from spot pricing. The thesis calculates "breakeven sizes" for each customer type such that larger customers should be on full spot pricing. These are on the order of one megawatt, for the customers and utility system modeled.

The thesis suggests that, for many large customers and independent power producers, conventional time-of-use rates are dominated by properly calculated prices which change every hour and are recalculated at least daily. For some other customers, rates of intermediate sophistication are best.

Thesis Supervisor: Richard Schmalensee
Title: Professor of Applied Economics

ACKNOWLEDGEMENTS

This thesis would not have been written without the assistance I received from a variety of sources. My advisers played key roles with their advice and support. Richard Schmalensee gave me over-all direction and provided incredibly detailed comments on my copious drafts. The reader should be grateful to him that this thesis is not even more wordy. Paul Joskow kept asking provocative questions, which led me into several fruitful areas. Fred Schweppe kept my analysis of electric power systems honest. He and Richard Tabors provided vital institutional background and contacts.

Michael Caramanis and Fred Schweppe helped me develop the material in Sections 3.1 through 3.3. It is theoretically plausible that each of us, working alone, could have produced the same final product. But it never would have happened.

At various stages I got useful comments on material from Jerome Delson of EPRI, Richard Gordon of Penn State, and Randy Ellis, Ben Golub, Nalin Kulatilaka, Mike Manove, Fred Pickel, and the Homeostatic Control Discussion Group, of MIT. Various seminar participants around the U.S. also gave helpful suggestions. Leigh Riddick and several utility people who must remain anonymous helped with the data used in Chapter 5. Alice Sanderson and Deborah Harrington did excellent typing of my many drafts.

I am grateful to the National Science Foundation for three years of financial support through its Graduate Fellowship program. This gave me the freedom to pursue my own research interests during the formative stages. The MIT Energy Laboratory provided useful office and support facilities for much of my time at MIT. The Sloan Foundation also provided some financial assistance.

Naturally, I alone am responsible for the remaining errors and omissions in this thesis.

Finally, I would like to dedicate my work to my wife, Liz. Her multifaceted support kept me going.

TABLE OF CONTENTS

	<u>Page</u>
Abstract	2
Acknowledgements	4
Table of Contents	6
List of Figures	10
Chapter 1 Introduction to Issues	13
1.1 Introduction to Issues: Electricity Pricing	16
1.2 Introduction to Issues: Other Markets	19
1.3 Main Contributions	21
Footnotes to Chapter 1	26
Chapter 2 Transactions Costs, Long-Term Contracts, and Optimally Sticky Prices	27
2.1 Explaining Long Term Contracts: Transactions Costs Opportunistic Behavior and Transactions Costs	32
2.2 Coping with Opportunism in Competitive Markets Determinants of Optimal Contract Duration	37
2.3 Application to Public Utilities	41
2.4 Long-Term Contracts and Risk Hedging	46
2.5 A Formal Model of Optimally Sticky Prices Problem Statement Alternate Interpretations Solving the Model Optimal Contract Duration Optimal Price Changing Indexed Prices	49
2.6 Some Implications for Pricing in Different Markets Application to Electricity	65
Footnotes to Chapter 2	71

Table of Contents, continued

	<u>Page</u>
Chapter 3	76
3.1	82
A Simple Model	
Model Formulation	
Model Solution	
Discussion of Basic Results	
Implications	
Optimal Investment	
Conclusion	
3.2	103
Extending the Basic Model	
Correct Full Spot Prices	
Spatial Pricing: The Two Area Case	
General Networks	
Rationing and Non-Spot Customers	
Decentralized Operation and Investment	
Conclusion	
3.3	120
Optimal Predetermined Prices	
Discussion	
3.4	125
Comparison with other Public Utility Models	
Time-of-Use Pricing	
Dynamic Pricing/Investment Models	
Spatial Pricing	
Pricing of Reliability	
Spot Pricing	
3.5	134
Assigning Participants to Different Rates	
Social and Private Optimal Assignment Criteria	
Implications	
Improving Self Assignment: The Adverse	
Selection Problem	
Mandatory Assignment	
3.6	146
Marginal Cost Pricing and Opportunism	
"Equal Revenue" Rates	
Conclusion	
Footnotes to Chapter 3	154
Appendix to Chapter 3	160

Table of Contents, continued

	<u>Page</u>
Chapter 4 Optimal Behavior Under Spot Prices Contributions of This Chapter Alternate Approaches	192
4.1 Basic Model Formal Statement of Problem	197
4.2 Examples Conclusion	203
4.3 Properties of Optimal Behavior Deterministic Behavior Value of Investments Stochastic Case Other Effects of Uncertainty Conclusion	209
4.4 Experience with Spot Prices	229
4.5 Conclusion Suitable Industries Market "Equilibrium" with Full Spot Pricing Application to Other Commodities	231
Footnotes to Chapter 4	240
Appendix to Chapter 4	247

Table of Contents, continued		<u>Page</u>
Chapter 5	Net Benefits of Spot Pricing: A Case Study	258
5.1	The Utility and Rates Modeled Estimating Full Spot Prices Rates Selected Prices Under Non-Spot Rates Simulating Forecasts of Spot Prices Conclusion	262
5.2	Case Study Customers Simple Storage Customer Discrete Rescheduling Customer Composite Shutdown Customer Standby Generator Customer	278
5.3	Results "Simple Storage Customer" Results Results for Other Customers Short Term Social Welfare Private Profit Under Different Rates The Value of Capital Investments	285
5.4	Transactions Costs, Customer Size, and Assignment to Rates Estimating Transactions Costs Customer Size and the Net Value of Spot Pricing	302
5.5	Changes in Utility Characteristics The "With Curtailment Premium" Case Results Value of Better Price Forecasts Generalizations	313
5.6	Nationwide Applicability of Spot Pricing Size Distribution of Large Customers Electricity Use by Suitable Processes Conclusion	324
5.7	Conclusions	334
	Footnotes to Chapter 5	338
	Bibliography	348

LIST OF FIGURES

	<u>Page</u>
2.0.1 Residential Electricity Price, Boston Edison Company	29
2.0.2 Possible Price Changing Patterns	31
2.5.1 Determination of p^* and L at time t	50
2.5.2 Sample Path of $p^*(t)$ Over Time	50
2.5.3 Hypothetical Pricing Policy and Realization of $p^*(t)$	53
2.5.4 Alternate Definitions of Variables	56
2.5.5 Sensitivity Analysis	63
2.6.1 Market Characteristics Leading to "Sticky" Prices	66
3.0.1 Outline of Chapter 3	80
3.1.1 System Short Run Marginal Cost Curve	90
3.1.2 Equilibrating Instantaneous Supply and Demand	93
3.1.3 Spot Price Over Time	98
3.1.4 Price Duration Curve	98
3.2.1 Issues Covered in Different Models	104
3.2.2 The Two Area Problem	108
3.5.1 Self Assignment Versus Socially Optimal Assignment	139
3.A.1 Simplified Electric Power Network	161
3.A.2 Private Versus Social Investment Criteria	185

List of Figures, continued

	<u>Page</u>
4.1.1 Multi Stage Production Process	198
4.3.1 Conditions at Optimum	212
4.3.2 "One Price Peak" Case	213
4.3.3 Pure Shutdown Case: Graphical Solution	214
4.3.4 Pure Storage Case: Graphical Solution	216
4.3.5 Mixed Storage/Shutdown Case: Graphical Solution	218
4.3.6 Two Peak Case: Graphical Solution	220
4.3.7 Non-Optimality of Certainty Equivalent Behavior	224
4.5.1 Characteristics of Processes	233
4.5.2 Characteristics of Prices	234
4.5.3 Full Spot Prices Before and After Implementation	237
4.5.4 Effects of Business Cycle on Net Revenues	237
4.A.1 Example of Optimal Charging Episode	255
4.A.2 Example of Optimal Discharging Episode	256
4.A.3 Example of Optimal Behavior	256
5.0.1 Case Study Calculation Flow	259
5.1.1 Actual System Lambda--August	264
5.1.2 Actual System Lambda--Selected Weeks	265
5.1.3 Price Duration Curve for 1980	266
5.1.4 Rates Used for Case Studies	268
5.1.5 Peak and Off-Peak Averages of System Lambda 1979-1980	274

List of Figures, continued

	<u>Page</u>	
5.1.6	Prices in August Under Two Rates	276
5.1.7	Summary of 1980 Prices Under Each Rate	277
5.2.1	Composite Shutdown Customer--Demand Curve	283
5.3.1	Results for Simple Storage Customer--Detail	286
5.3.2	Social and Private Savings Due to Time Varying Prices	288
5.3.3	Results for Simple Storage Customer	290
5.3.4	Results for Discrete Rescheduling Customer	291
5.3.5	Results for Discrete Composite Pure Shutdown Customer	292
5.3.6	Gross Social Savings Under Each Rate	294
5.3.7	Total Value of an Investment in Storage Capacity	300
5.3.8	Marginal Value of Production Capacity	301
5.4.1	Summary of Estimated Transactions Costs	307
5.4.2	Net Social Savings as a Function of Customer Size and Rate	310
5.4.3	Socially Optimal Assignment to Rates, by Customer Size and Type	311
5.5.1	Full Spot Prices: Base Case Versus "With Curtailment Premium" Case	315
5.5.2	Results for "Spot Prices with Curtailment Premium" Cases	315
5.5.3	Gross Social Savings--Sensitivity Analysis	316
5.5.4	Social Value of Better Spot Price Forecasts	320
5.6.1	Size Distribution of Customers	326
5.6.2	Comparisons of Estimated Energy Use by Large Customers	345
5.6.3	Electricity Use in Suitable Processes	329

CHAPTER ONE
INTRODUCTION TO ISSUES

This thesis is about the possibility of spot pricing public utility services. Spot prices are prices which vary over time in response to current conditions. Spot pricing of utility services was proposed by Vickrey in 1971, but until recently has been essentially ignored.

Traditionally, electricity and other public utilities in the U.S. have been sold to retail customers at a constant price from hour to hour, and month to month. In fact until the adoption of automatic adjustment charges in the 1970's, electricity prices were often constant for a year or more. In about 1960, Boiteux introduced the concept of "time-of-use" prices; prices which varied systematically from hour to hour, according to a set, predetermined pattern. Recently this has begun to be adopted in the U.S., especially for large electricity customers. It has been the norm in the U.S. for long distance telephone calls and in Western Europe for electricity. "Time-of-use" prices typically divide the day into two to four periods, with a different price per unit in each period. Each weekday repeats the same cycle; weekends have prices of their own, or have the lowest weekday price.

Spot pricing goes systematically beyond time of use prices. The basic idea is that instead of charging a price which is predetermined up to a year in advance, the actual price at any instant is calculated

only a moment before it takes effect. For example in "five minute spot pricing" a new price is calculated each five minutes and is in effect for the next five minutes. Since the price at each moment is based on current conditions, the pattern of prices over a day may look quite different than the pattern of prices one or seven days earlier.

In fact the distinction between "predetermined" and "spot" prices is somewhat gray, since no prices are fixed forever, and conversely it is not conceivable to change the price each passing instant of time. Thus any real pricing scheme must have some degree of predetermination in it. This concept of a continuum of pricing methods will be developed formally and exploited in Chapters 2 and 3.

But in practise there are two very different philosophies at work. The analysis of conventional predetermined prices emphasizes setting the best possible prices fixed for the indefinite future. As conditions unfold from day to day prices are held constant, and other factors, especially the production pattern of different units, are varied to adapt. In extreme circumstances total demand for electricity may exceed total supply, and rationing of various kinds will be implemented. In contrast, the analysis of spot prices emphasizes using the current price as an active control instrument in a feedback loop. The current price reflects current conditions, and is one method of helping the utility system adapt to changing conditions. If spot pricing is fully implemented, involuntary rationing of customers is theoretically never necessary, as supply-demand imbalances can be removed by means of price adjustments.

In theory, spot pricing can provide a substantial efficiency improvement over predetermined prices. For electricity, average generating costs are reduced, customer and utility profits are improved, and blackouts and other disruptions are reduced or removed entirely. However to implement full spot pricing, a more elaborate system is required for calculating and communicating prices and metering the resulting customer behavior. Various transactions costs are increased over those required for predetermined prices. These increases may more than compensate for the theoretical efficiency improvements provided by full spot pricing. Therefore forms of pricing intermediate between traditional and full spot pricing may be better than either extreme. One purpose of this thesis is to begin a quantitative cost/benefit comparison of different pricing systems.

The rest of this chapter formally presents the issue of price changes over time, first for electricity (Section 1.1), then for markets in general (Section 1.2). Section 1.3 discusses the main contributions of the thesis.

1.1 Introduction to Issues: Electricity Pricing

As conditions change, so should prices. This obvious concept has not been subjected to systematic study. Economic models have typically looked at a single static equilibrium, or assumed instantaneous adjustment to a series of static equilibria. When conditions are changing slowly or adjustment costs are low, this is a good approximation. But it is inadequate when they are changing rapidly.

The economics of price change have probably been studied as intensely for electricity and public utilities in general as for any other industries. This is appropriate for several reasons. First, conditions in electric power systems change very rapidly. For example, demand changes of 30% over 12 hours are common. And the magnitude of hour to hour changes of conditions in electricity markets has recently increased. (See Section 5.5.) Therefore errors caused by analyzing conditions as if they were in static equilibrium are larger now than a few years ago. Second, public utilities are subjected to price regulation and deliberately isolated from market discipline. Therefore pricing methods must be chosen based on intellectual understanding, rather than reliance on "blind" market forces. Third, these industries are quite important in the U.S. and other economies.

Despite this importance, the main stream of literature on public utility pricing has embodied several implicit assumptions which

severely limit its validity and applicability in current conditions. The basic thrust of this literature (which is variously referred to as "time-of-day", "time-of-use", or "peak load" pricing literature) is to analyze optimal prices as a function of predictable variations in demand. The simplest assumption, used in the early studies, is that demands follow a deterministic repeating cycle, year in and year out. This assumption does give useful insights into how prices should change, but it is too simple for practical application because actual demands are stochastic. Therefore more recent models have treated demand as following a cycle with a stochastic component. But the probability distribution of demand is assumed to be known in advance and unchanging from year to year. Therefore once appropriate prices are set, they remain fixed from week to week and year to year.

Of course a possible defense of such models is that as demand and other parameters change from year to year the models can just be re-solved with the new values to get the new optimal prices. This is a "successive static equilibrium" approach once again. It does not help with the following important questions:

- o How often should new prices be chosen based on the new demand and supply parameters? Once a year? More or less often?
- o What is the impact of uncertainty about the parameters on the optimal current prices? On optimal current investments?

- o How long should the pricing periods within each cycle be, and how should they be chosen? In practise time-of-use prices have divided the day into several periods of two to fourteen hours each, but this has been done without formal analysis. Theoretical models have derived formulas for pricing over N periods within a cycle, but not analyzed how N should be chosen.
- o How do time-of-use or other time varying prices interact with rate of return regulation? If regulators adjust marginal cost based time-of-use prices next year based on what happened this year, that may give the regulated utility an incentive to deliberately raise its marginal costs. How does this affect "optimal" prices?

This thesis considers the first three of these questions thoroughly, and the last one briefly. The discussion and models focus on electricity. Other utilities may have different technological constraints and operate on slower time scales. But the same qualitative insights apply.

1.2 Introduction to Issues: Other Markets

Obviously the problem of how to change prices over time is not confined to electricity or other regulated industries. Few markets have the strong diurnal demand fluctuations that electricity does, but all are characterized by some stochastic variations in both demand and production costs. Hence the conventional static equilibrium model used to describe many markets is only an approximation. Nor can most markets be characterized as a succession of perfect static equilibria. Therefore the analysis of this thesis, which emphasizes electricity, is to some extent useful for understanding the behavior of many unregulated markets.

Unregulated markets show various patterns of price changes over time. Most market exchanges between firms take place between the same firms over a period. At the beginning of any such long term bilateral relationship, the parties must explicitly or implicitly make several agreements. These include the specifications of the product being sold, the price per unit, and the quantity to be sold. Typically, in long term contracts the purchaser can specify the quantity each period (within limits), and pays a price per unit which may change over time. The price can be adjusted from the initial price by three methods: prespecified changes at prespecified times, adjustments based on selected variables which are not known when the initial contract is signed (indices), and by an agreement to simply renegotiate the whole contract at a prespecified time.

Why are different price adjustment methods used in different bilateral exchanges? As I will show in Chapter 2, the key tradeoff is between the transactions costs of changing prices more often, and the profit/welfare losses due to changing them less often. The transactions costs depend on the market structure and the nature of the product, for example the number of producers and users of the product. The production cost increases due to infrequent price changes depend on the production and use technologies for the product, and on the exogenous factors affecting producers and users (such as prices of inputs and demand for the final product produced by the user).

The overall structure of the market is also relevant. For example, a competitive market will behave differently than one controlled by a monopolist. A regulated monopolist will behave differently than a profit maximizer or a welfare maximizer. In each case the relevant decision makers will trade transactions costs against potential profits but the outcomes may be very different. For example price changes may threaten the stability of an oligopoly, and this will cause higher implicit transactions costs for negotiating and implementing them. Therefore price changes will probably be less frequent in oligopolies than in competitive markets, and deviations from the oligopolists' "optimal" prices in a static model may be large.

Different markets will have different patterns of price change. By analyzing the structure of transactions costs, technology, and the exogenous forces on the market, it is possible to predict how prices will adjust over time. This is done at a qualitative level in Chapter 2. Subsequent chapters do this in more detail for electricity.

1.3 Main Contributions

This thesis shows how to describe different systems for changing prices, and how to evaluate their relative merits. It conducts such analyses only to a limited extent and only for electricity. Nonetheless its results suggest that current electricity pricing practises in the U.S. (and probably elsewhere) are not maximizing total social welfare. They can be changed in directions which will increase the welfare of both customers and electric utilities.

The thesis makes seven main contributions to previous analyses of public utility and other pricing.

1. It gives a general specification of possible mechanisms for changing prices over time. Previous analyses have assumed a price changing policy without making sure it was optimal. (Chapter 2)

2. It shows how to calculate socially optimal electricity prices and investments when transactions costs are zero. The model developed is considerably more general than anything in the past, as it includes:

- o the transmission and distribution network; spatial pricing
- o pricing methods varying from "pure spot" to "completely predetermined"
- o the price responsiveness of demand
- o different pricing methods for different customers

- o intertemporal relationships of demand
- o intertemporal relationships of generation costs
- o generation by independent firms (those not centrally dispatched)
- o investment by customers and the central utility.

All of these except investment are modeled as stochastic. Past engineering and economic models of electric power systems have considered only a few of these issues at a time.^{3.1} Stochastic outages of generation, which are extremely important in real electric power systems, have been virtually ignored in the past.

The derivation of optimal prices developed here is quite simple. Previous formulas for optimal time-of-use prices are easily derived as special cases. The formulas for optimal investment under different pricing methods are even more straightforward, in contrast with the convoluted derivations necessary in past approaches. (Chapter 3)^{3.2}

This model of optimal pricing can be interpreted as a method for deregulating electric generation, with a central transmission utility which sets buying and selling prices based on the rules derived here. This concept will not be extensively discussed; see Bohn et al [1982].

Another implication of this analysis is that firm level economies of scale for electricity have been improperly analyzed in the past. Whether or not there are such economies depends on the existence, nature, and transactions costs of spot markets between

utilities and generators. The relevant policy issue is whether horizontal integration, or markets, leads to lower total costs. This has not been addressed in past econometric studies of scale economies.

3. It provides a qualitative discussion and simple formal analysis of the effects of transactions costs on optimal pricing. This model is developed for the general case. (Chapter 2)

4. It analyzes how different customers should be assigned to different pricing methods. The possibility of letting customers self select is discussed. (Section 3.5)

5. It discusses the conflict between marginal cost pricing and incentives for efficient operation by regulated utilities. While this conflict cannot be eliminated, its strength can be made essentially invariant to the price adjustment method used. (Section 3.6)

6. It develops several models of customer behavior in response to stochastic, time varying prices. These mathematical models are designed for the analysis of large customers for whom econometric analysis is inappropriate or impossible due to lack of data. Past optimizing models of customer response to prices have assumed deterministic prices and have been primarily ad hoc. (Chapter 4)

7. It provides a quantitative estimate of the benefits of different pricing policies for electricity for selected "case study" customers. The benefits are calculated as functions of utility characteristics, customer characteristics, and pricing methods. The results suggest that for large industrial customers with certain

technologies, conventional time-of-use prices give social costs significantly higher than do optimal prices. Since large industrial customers are a significant fraction of demand for most utilities, this has strong practical implications, especially when utility opportunism will not be affected by the pricing method used.^{3.3} (Chapter 5)

This thesis emphasizes prices which differ across states of nature, i.e. in response to the revelation of uncertainty. It also discusses how to set prices across space and time. Pricing across space and time is mathematically similar to pricing across states of nature, to the extent that behavior at each moment or location is independent of behavior at other moments or locations. This means that much of the earlier work on time-of-use pricing in deterministic models carries over to the theory of spot pricing. For example, the effects on optimal generating mix of going from flat to predetermined time-of-use prices are analagous to those of going from time-of-use to spot prices. Such relationships facilitate analysis and understanding.

Nonetheless deterministic time-of-use models are completely inadequate to fully understand spot pricing. The practical issues of how to implement and evaluate spot prices are very different than those for time-of-use prices. For example, engineers and economists have spent a lot of time developing non-price methods (direct load control and rationing) of solving uncertainty induced problems, even when they accept time-of-use pricing.^{3.4} Similarly, conventional wisdom seems to be that it will be much harder to get customer acceptance of spot pricing than of time-of-use pricing. Current time-of-use pricing experiments

yield very little information about how customers would respond to spot prices.

On a more theoretical level, uncertainty raises some issues not present in time-of-use models. Many of these have to do with the value and timing of information. Transactions costs due to potential opportunism are a key issue not present in deterministic models. (See Chapter 2.) Determining optimal behavior by generators and users is harder and more interesting under uncertainty. (See Chapter 4.) How to value generation by customers who sell "unreliable" amounts of power back to their utility has been a very controversial question, but is easily answered in the fundamentally stochastic approach of spot pricing. Therefore despite the mathematical similarities between spot pricing and previously analyzed time-of-use pricing, this thesis provides new insights into public utility pricing.

FOOTNOTES TO CHAPTER 1

3.1 The closest economic models to the one developed here are perhaps the intertemporal/spatial models of Takayama and Judge [1970]. However their models were deterministic, were restricted to linear supply and demand functions for simplicity, were not directly applicable to electricity, and were not interpreted as leading to systematic policies for changing prices over time. See Section 3.4.

3.2 Almost all the material in this category was developed jointly with Michael Caramanis and Fred C. Schweppe. See Bohn et al [1981] and Caramanis et al [1982].

3.3 For example the 500 industrial customers placed under mandatory time-of-use pricing in one large Wisconsin utility's territory account for 37% of the energy it sold. [Malko and Faruqui, 1980]

3.4 Some of these proposals are covered in Section 3.4. Some economists who have written on public utility pricing have rejected spot pricing outright.

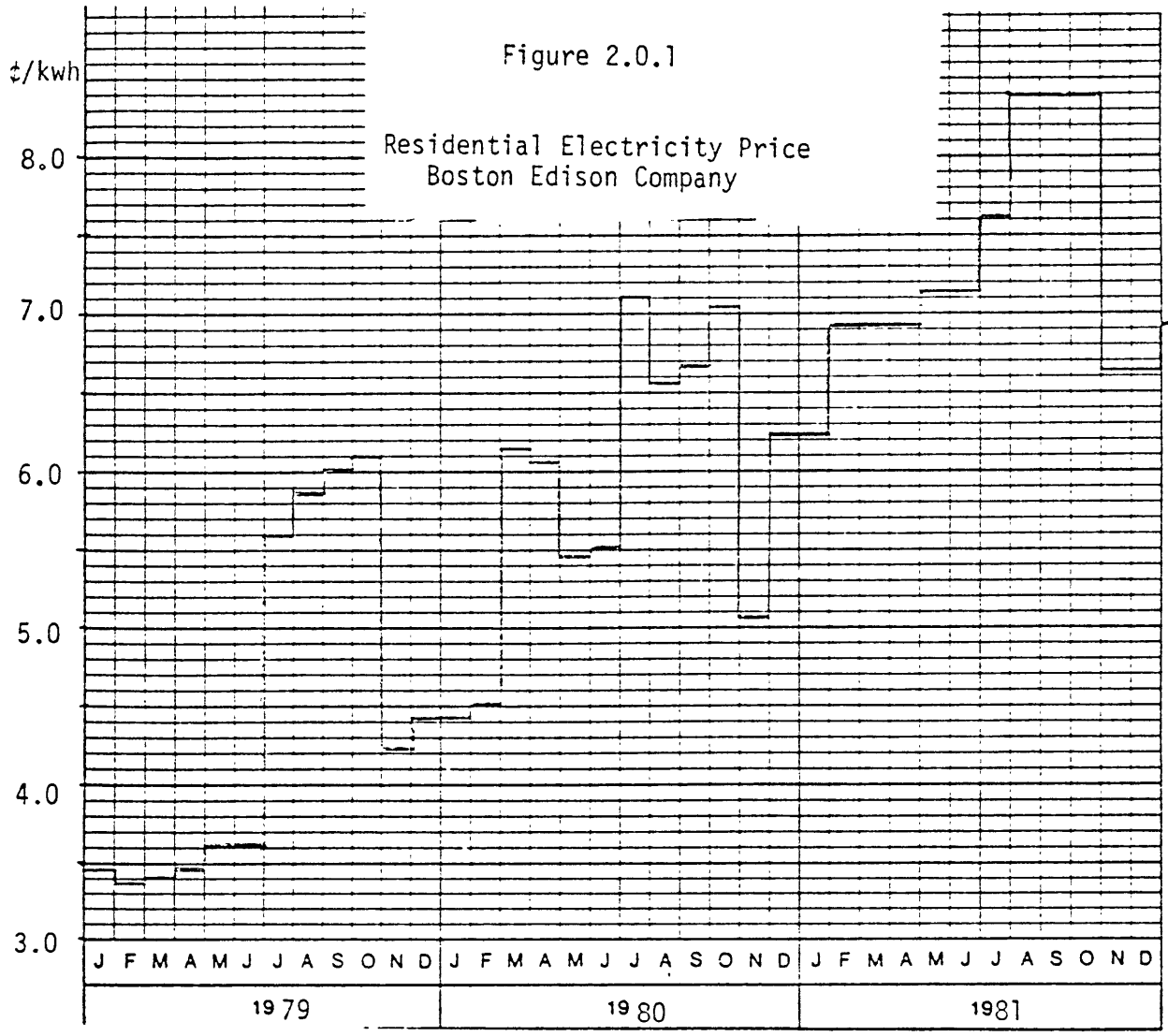
CHAPTER TWO

TRANSACTIONS COSTS, LONG-TERM CONTRACTS, AND OPTIMALLY STICKY PRICES

Standard economic models predict that exchanges over time will be a series of discrete instantaneous spot market transactions, with prices continually adjusting to reflect marginal conditions. But in most real markets, specific buyers and sellers deal with each other over long periods, at prices which change at intervals and in discrete jumps. This chapter shows when, why, and to what extent market participants prefer long term contracts with discrete price changes, over continuously adjusting prices. The analysis centers on reducing transactions costs, and develops a framework for the design of pricing procedures which incorporate transaction costs. Competitive markets are considered first, then public utilities. Later chapters use this framework for an analysis of electricity pricing.^{0.1}

Anecdotal knowledge shows that prices do not change continuously, although data on the frequency and magnitude of actual price changes for specific products is limited. Stigler and Kindahl [1970] undertook an extensive survey covering data from 1957 to 1966. Carlton [1981] analyzed their data further and showed that price rigidity (interval between price changes) differs widely both within and between different industries. Gordon [1975, Chapter 3] examined coal purchase contracts in the U.S. and found widespread use of both short (less than one year) and long (greater than ten years) purchase contracts.^{0.2} Consumer goods show conspicuously different amounts

of price rigidity. (Compare mail order goods, restaurant prices, and supermarket prices.) Finally, historical price data is publicly available for many regulated industries and shows varying patterns of price rigidity. Figure 2.0.1 shows the electricity price for residential customers of Boston Edison over several years. Notice that these prices fluctuate greatly from month to month.

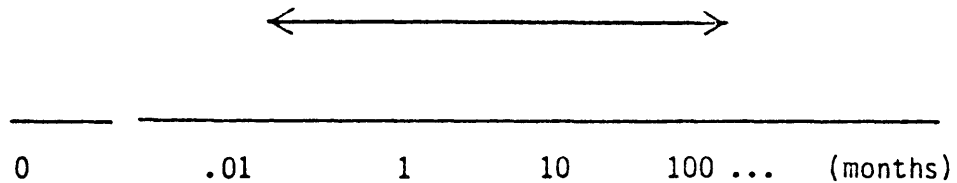


Source: Price per kilowatt hour for all use beyond 1000 kwh/month.
Includes fuel adjustment charge and summer surcharge.
From Boston Edison Company bills

Why do prices change in discrete jumps, instead of continuously as predicted by standard economic models which determine the optimal level of prices by looking at successions of static equilibria? Why are there such differences in price rigidity among different products? The basic reason is that various costs are incurred by the act of changing a price. Any dynamic truly optimal pricing rule must consider these costs, and weigh them against the benefits of changing prices. Such trade-offs are made in competitive and monopolistic markets, and should be made in markets with regulated prices. Hence markets have "sticky" prices. That is, the actual price changes less often than the underlying "optimal" spot price predicted by a conventional model.^{0.3}

Why not go all the way to permanently fixed prices? What are the costs of leaving a price fixed (in nominal terms) over time? They are the losses due to deviations between current price and the "optimal" spot price as measured by a model which ignores transactions costs. For example, if marginal production costs rise but a monopolist keeps its price the same, it sacrifices some potential net revenues. But it avoids the costs of a price change. The monopolist can determine its optimal system for changing prices by balancing the expected value of lost net revenues against the transactions costs of different price changing policies. Figure 2.0.2 indicates the spectrum of price changing policies available, from continuous spot prices to fixed prices guaranteed in perpetuity.^{0.4} The ideal pricing policy in any market will be a function of the volatility of the environment, the size of customers, and the costs of price changes.

Figure 2.0.2
Possible Price Changing Patterns
(Mean Interval Between Price Changes)



This chapter examines these issues for markets in general. Section 2.1 presents the key paradigm: long term contracts, which wholly or partially fix prices in advance, reduce transactions costs. The most important issue in many markets is the role of such contracts in reducing transactions costs, notably the costs of opportunistic behavior. Section 2.2 details the variety of contractual techniques used in competitive markets to reduce opportunistic behavior. Section 2.3 shows that different methods of public utility regulation can be interpreted in terms of these techniques. Section 2.4 discusses the role of risk sharing in explaining long term contracts and therefore in explaining sticky prices. It argues that long term contracts designed to redistribute risks are qualitatively different than those designed to reduce transactions costs.

Section 2.5 provides a simple analytical model of some of the ideas in the chapter. Optimal price changing patterns are derived as a function of various market parameters. The model illustrates the key issues, which are elucidated and quantified for electricity in

later chapters. For example, it shows that while the predictable component of changes in exogenous conditions can be dealt with by predetermined changes in actual prices, the random component of changes in conditions can be handled only by spot (adaptive) price fluctuations. Thus time-of-use public utility prices, which are predetermined, have only limited effectiveness. They may nonetheless be preferable to full spot pricing, because of lower transactions costs.

Section 2.6 summarizes the types of markets in which prices should be quite "sticky". That is, they will often deviate from the underlying optimal spot prices. This summary predicts the comparative level of price stickiness in various unregulated markets, and it should be useful for choosing the level of price stickiness in regulated (public utility) markets. Some normative implications for electricity pricing are discussed.

My analysis owes a heavy debt to Williamson, particularly his 1979 article which develops the central importance of transactions costs in determining contract form. This chapter attempts to further develop the relationships between transactions costs and predictable and unpredictable change, and to apply this to public utility pricing.

2.1 Explaining long term contracts and sticky prices: Transactions Costs

Most repetitive purchases of a product by one firm take place within an on-going relationship with another firm, rather than through the classic anonymous spot market. Typically the two firms have a

formal or informal relationship which pre-specifies some terms of the purchases, especially the price per unit. I will refer to these as long term contracts. They are the crucial focal point for the analysis of sticky prices and transactions costs.

Given that two firms have entered into a long term contract to buy and sell a good, what form will it take. Standard micro theory suggests the contract will allow the terms of the sale to vary continuously as conditions vary. Specifically, as conditions in upstream and downstream markets change, the marginal cost of production and the marginal value of the good to the user will shift, leading to a shifting optimal spot price.^{1.1} Long term contracts with sticky prices thus appear to be Pareto inferior; buyer and seller could each do better by letting the actual price equal the optimal spot price at each moment, and agreeing in advance to side payments to redistribute joint profits. Such contracts would Pareto maximize each firm's gross revenues minus direct production costs.^{1.2}

This analysis, however, neglects transactions costs. A long term contract which allows for continually shifting prices will have high costs for calculating the price at each moment, measuring usage at that price, and otherwise monitoring the situation. If side payments must be calculated, transactions costs will be even higher. Thus the objective of each firm is to maximize the net present value of: gross revenues minus production costs minus transactions costs. Since transactions costs are an increasing function of how often prices change, including them in the objective function will lead the two

firms to agree to somewhat "sticky" prices. How sticky will depend on the relative effects of price changes on transactions costs and production costs.

This intuitive argument will be formalized in Section 2.5. For now I turn to the determinants of transactions costs. In most situations, the mundane costs of metering and billing are only a fraction of the relevant transactions costs.

Opportunistic Behavior and Transactions Costs

Consider a long term contract between two firms, a supplier and a buyer. Potentially the largest component of the transactions costs of this contract, and one which has a critical influence on the form of the long term agreement the two make, is opportunistic exploitation by each firm of idiosyncratic investments by the other. [Williamson, 1979] An idiosyncratic investment is one which has more value in that specific exchange relationship than in any other. The extent to which an investment is idiosyncratic depends on the availability of close substitutes and of alternate suppliers and buyers for the product. For example suppose the supplier sells an integrated circuit (IC) to the buyer. A custom IC may give the buyer's product better performance or lower total cost by reducing the total circuitry needed. But when the buyer seeks such savings by designing its product around the custom IC, it is making an idiosyncratic investment. Conversely the manufacturer has invested in designing and manufacturing the IC. These constitute investments in specialized

capital, with salvage value less than their original cost. Both buyer and seller are then "locked in" to their relationship to the extent that the investments are idiosyncratic to their relationship. (More precisely, they are locked in to the extent of the "appropriable quasi rents" in their investments. See Klein et al. [1978] and Williamson, [1979, footnote 30].)

In effect, an initial arm's length relationship between potential buyer and potential seller is now a continuing relationship of partial bilateral monopoly, arising out of the idiosyncratic nature of the product and the relationship. One or both parties are vulnerable to opportunistic demands by the other for more favorable prices on the product. This is an inevitable outgrowth of making a specialized investment.

Therefore when contemplating such investments, each party will attempt to minimize the other's power to expropriate some of the quasi rents once its investment is made. There are many ways to do this, corresponding to different contract types. The most obvious approach is a rigid long-term contract for the life of all specialized investments. Both price and quantity to be exchanged would be fixed in advance, so that neither buyer nor seller is exposed to opportunistic attempts by the other to extract quasi rents.

But this method of ensuring against opportunism does so at great cost in flexibility, which also has economic value. Both buyer and seller would like to be free to adjust their purchases/sales and the price as conditions change. Ideally, they would like to follow the

"full spot" prices, perhaps with side payments. For example if demand for other IC's increases and IC production equipment is running at full capacity, the opportunity cost to the seller of producing for this buyer increases. Therefore buyer and seller will attempt to negotiate a contract which maintains flexibility for both while minimizing their exposure to opportunism. The type of buyer/seller relationship ultimately chosen (from successive spot sales at one extreme, to prices fixed in perpetuity or vertical integration at the other) reflects the trade-offs among their desires to:

1. Encourage specialized investments, which lower joint production costs.
2. Discourage the resulting potential opportunism.
3. Allow for flexible, joint production cost minimizing prices and quantities. The optimum optimorum for this purpose is the classical static equilibrium price/quantity combination at each moment. (I will refer to these prices as the "full spot" prices.) One way to reach it is to set the price at the full spot price each moment and let the buyer choose the quantity it purchases. Thus we can restate the problems as "minimizing deviations from full spot prices". The metric used to measure "minimizing" will be discussed later; ordinary least squares is a convenient approximation.^{1.3,1.4}
4. Minimize the transactions costs of negotiating, monitoring, and enforcing an agreement which does all of the above.

The first and third objectives are concerned with maximizing total profits (buyer and seller of the product). The second and to some extent fourth are concerned with the jockeying by the two parties to increase the profit of one at the expense of the other.

2.2 Coping with Opportunism in Competitive Markets

What mechanisms are used in competitive markets to cope with potential opportunistic behavior while maintaining flexibility and minimizing deviations from optimal spot prices ^{2.1}

There are several ways to allow flexibility in long-term relationships while discouraging opportunism, when faced with uncertainty and bounded rationality. Ideally these mechanism are agreed to by both parties, before either makes significant idiosyncratic investments. They may all lead over time to deviations from optimal spot prices. All of them are potentially useful for regulating public utilities, as discussed in the next section.^{2.2}

1. Contingent contracts. Specify in advance a large number of contingencies (strikes, changes in input and output prices, etc.) and what should be done to the price and quantity for the product in each case. Complete specification of all contingencies is infeasible because of bounded rationality; and attempts in this direction have very high transactions costs for initial negotiations and continual monitoring ("What really did happen").
2. Allow quantity adjustments (within a band) at the buyer's initiative, but at the original price. If the buyer wants to

reduce its purchases but is not turning to another supplier, it is unlikely to be exploiting the seller.^{2.3} This is better than not allowing any adjustment; but over time the optimal spot price may deviate greatly from the original price. Since the selling price is not allowed to change, this leads to production inefficiency.

3. Allow quantity adjustments at a price which changes via escalator clauses. These escalator clauses should be based on objectively verifiable and exogenous events (such as prices of raw materials). Otherwise opportunistic manipulation of the escalator clause is possible. The ideal case is when a closely-related product is traded in an independent spot market which can be used as an index. For example if a producer sells in two markets, with different degrees of market power, buyers in one market might negotiate an escalator tied to the price in the other market. This will lead to prices which inefficiently mirror demand conditions in the second market, but they will at least fairly reflect changes in producer costs. Another example is the use of predetermined price changes (x percent per year, for example). The index is calendar time, which is certainly exogenous and cheaply verifiable. Predetermined price changes are modeled in Section 2.5.
4. Devise methods which remove the incentive for opportunism, if not the opportunity. Explicitly monitoring the other party's profits and limiting them to an agreed-on amount is one approach,

commonly used for regulated industries. Such cost plus fixed fee contracts, however, weaken incentives for efficient production and use.^{2.4}

5. Rely on an outsider to arbitrate price changes. Williamson refers to this as "trilateral governance" (p. 249) and suggests it is most useful for products which are traded only occasionally and thus do not justify the expense of developing a special governance/contract structure. (Public utility regulation does not fit in this category; see Section 2.3.) The third party arbitrates unexpected situations.
6. Rely on the market for discipline. Make all transactions through spot purchases and sales. This avoids bilateral relations entirely, hence avoids opportunism only by making sure that specialized investments are not undertaken. It is therefore useful mainly when the product involved is "not idiosyncratic"--that is, total demand is large enough to support multiple buyers and sellers. Otherwise, it sacrifices production efficiency. Notice that product selection becomes endogenous. Buyers and sellers can rely more on a market for price discipline by using or making a more standard product [Carlton, 1981; Wachter and Williamson 1978, page 564].
7. Withdraw a transaction from the marketplace entirely, by vertically integrating. This sacrifices scale economies in production, which severely limits its applicability to most public utilities.

I have discussed these mechanisms as if they were for bilateral relationships between one buyer and one seller. Most of them can also be applied to transactions between many buyers and one seller. Using a single "buyer's agent" to conduct one side of the negotiations will reduce transactions costs. And some of the mechanisms may lead to more efficient prices if they are implemented for many customers in a single agreement.^{2.5}

Determinants of Optimal Contract Duration

In all the above cases except the last two, contract duration must be mutually chosen. Each time a contract expires it must be renegotiated, and there is a chance that the other party will bargain opportunistically for more favorable renewal terms. Therefore any party considering a potentially appropriable investment in an idiosyncratic product will want a contract duration long enough to recover the appropriable portion of its investment. Thus the larger or more specialized the investment, the longer the contract duration. On the other hand higher uncertainty about the future will give both parties an incentive to push for shorter contract duration or more comprehensive (and opportunism resistant) adjustment mechanisms. These mechanisms may include more frequent price recalculations. During the 1970s the rate of inflation increased as did uncertainty about the future rate of inflation/ cost escalation. The above theory predicts that this led to shorter contract durations and more frequent price recalculations. This will be shown formally in Section 2.5.

2.3 Application to Public Utilities

The above analysis also applies to public utilities. Consider a privately owned company which is awarded a monopoly franchise to provide a service. It has the potential to opportunistically exploit the idiosyncratic portion of investments made by potential customers within its franchise territory. Their mere decision to locate within its territory is such an investment. Therefore, as a condition of awarding the franchise, customers (acting through their agent, a government) will require some protection against potential exploitation. This takes the form of rules governing the utility's pricing and other behavior. The private firm, in turn, faces potential exploitation by the buyer's agent; therefore it will demand some long term protection before making idiosyncratic investments in the franchise.^{3.1} Thus public utilities are a special case of Williamson's model of recurrent bilateral exchange involving idiosyncratic investments.

What provisions are used to adjust public utility prices In electricity, three main mechanisms have been used: holding prices essentially fixed; removing the incentives for opportunism; and using supposedly exogenous price indices for price adjustments. Other methods more suited to current conditions may be available and are discussed.

Traditional regulation: Pre-1970 electricity regulation basically followed the second method above: customers were allowed to choose the quantity (with an upper bound) while prices remained constant or

were unilaterally lowered by the producer/utility. [See Joskow 1974] This worked when uncertainty was low and price flexibility was not important. It broke down when conditions changed, and especially when the rate of change and the level of uncertainty became too high to be handled by existing mechanisms. Using the above terminology, the constant prices led to large deviations from optimal spot prices, hence to large increases in total production costs and to significant redistribution of profits from producers to users.

Current regulation: Current regulatory practise can be characterized as the fourth method of Section 2.2: remove, as far as possible, the incentives for opportunistic behavior. This is done by setting revenues equal to out-of-pocket expenses plus a more or less fixed fee tied to the level of investment (rate of return regulation). The well known problem with this approach is that it controls the incentives for opportunistic price manipulation, but creates incentives for socially sub-optimal investment and operating behavior. It also leads to inefficient incentives for customer behavior, due to still large deviations from full spot prices.

Within the current approach, two adjustment mechanisms are used: formal regulatory hearings, and automatic adjustment clauses. Formal hearings have very high "negotiating" costs and are very slow, making large deviations inevitable. Therefore automatic adjustment clauses are being used increasingly. However these have not been the exogenously based adjustments cited (Method 3) above. They are

instead based on variables within the firm's control. For example total costs of fuel purchased are used as an index, rather than an exogenous price per gallon in another market. This may bias the utility toward fuel intensive technology. More complete adjustment clauses reduce this bias, but also remove even more incentives for cost minimizing behavior (X efficiency [Leibenstein, 1980]).

Schmalensee [1979, p 109ff.] discusses the trade-off between input bias and X-inefficiency in cost based adjustment clauses.

Time-of-use prices: There is now increasing use of a form of the second adjustment mechanism: price adjustments based on truly exogenous variables. The variables chosen have been time of day and season. Such time-of-use prices have been favored by economists because they can more closely track optimal spot prices (which change cyclically by day and season). The economic literature has favored calculating these prices based on the utility's marginal costs.^{3.2}

Unfortunately, practical marginal cost based rates do not avoid the conflict between opportunism and efficient price adjustment over time. It is essentially impossible to measure the marginal cost without using variables which are somewhat under the control of a monopolistic utility. Thus the utility has the ability to opportunistically affect the price charged. (The utility has four sets of control variables for this purpose: dispatch order, plant outages, fuel purchases, and plant construction. A simple example is that the utility would profit from reduced peaking capacity, since this would raise the frequency of high marginal costs.)

Any time-of-use prices which are "optimal" in the static and deterministic sense of the traditional economic literature on public utility pricing have the same problem. Essentially, such prices are weighted expectations of the optimal spot prices. Hence they are affected by everything which affects the optimal spot price. See Section 3.3. So as long as the utility can influence the way marginal costs are calculated, prices based only on marginal costs will give an opportunity and an incentive for opportunistic behavior.^{3.3}

Merging These Approaches: Under current regulatory practise, the utility can also influence its prices, but the approach tries to remove incentives to do so by making profits essentially independent of utility behavior. This same mechanism can be used with flat, time-of-use, or optimal spot prices. A "revolving fund" or other ex post mechanism would be used to make sure that the utility effectively rebated all revenues in excess of out-of-pocket expenses plus a constant.^{3.4}

The revenues of such a revolving fund could be rebated selectively to different customer classes, based on some relatively load independent measure of customer size, or pro-rated across all kilowatt hours.^{3.5} Of course not all prices could exactly equal optimal spot prices in this situation. This approach will be discussed in detail in Section 3.5.

Increased Competition: Still another potential mechanism for reconciling static efficiency with reduced opportunism is number six above, i.e. using the market for discipline. Short of major deregulation, it is still possible to encourage competition by utilities for the business of large customers cited near boundaries. Moves to mandate "wheeling" of power from independent generators are of this type. Not only does this reduce potential conventional static loss due to monopoly, but it provides the customer more flexibility as conditions change. And unless the customer is very large relative to the total utility (which is extremely rare) it provides the customer no significant monopsony power hence cannot be objected to on those grounds.

A fully competitive market for electricity might also be feasible, if properly designed [Bohn et al., 1982]. But technological constraints may require that a monopolistic public utility own most transmission facilities. Therefore even with a deregulated generation market, all the above trade-offs still apply to regulating the mark-up charged by this utility. That is, the utility buys from suppliers at $p(t) - m(t)$ and sells to users at $p(t)$, where $m(t)$ is the markup or transportation charge. There is a "socially optimal spot markup" for $m(t)$, as detailed in Chapter 3. The problem is to regulate the transmission utility so that it actually charges this $m(t)$, without giving it perverse incentives to mis-invest or otherwise misbehave. This is almost exactly analogous to the problem of regulating $p(t)$ for a utility which owns its own generation.

2.4 Long-Term Contracts and Risk Hedging

So far we have considered long term contracts solely as a vehicle for trading flexibility (and hence joint production costs) against transactions costs. They also have another role: hedging against price fluctuations.^{4.1} Fortunately, and contrary to some intuition, these roles do not conflict: hedging can be achieved without affecting the production cost/transactions costs trade-off. The reason is that for production efficiency, only the marginal price is relevant; while for risk hedging, only the average price is relevant. Because the subject comes up frequently when discussing long term contracts, I develop this point here.

In a standard bilateral forward contract, both price and quantity to be exchanged at a specified time are completely pre-specified, providing a means of completely hedging against the risk of fluctuations in the current price. [Black, 1976] Such contracts do not provide for any production flexibility, and therefore cannot be the only method of exchanging a good. A market based solely on such contracts would soon find itself in the situation of a centrally planned economy, with too much or too little of the product being produced. For non-idiosyncratic goods (pure commodities), pure spot markets provide flexibility; this is the case usually considered in financial and economic models. But as discussed, for most goods spot transactions will not provide the optimal trade-off of production and transactions costs.

Therefore, firms which want price hedging for idiosyncratic

products must use forward contracts in conjunction with one of the flexible quantity contract arrangements of the type discussed in Section 2.2. Suppose that buyer and seller have worked out the optimal long term contract considering only the production cost/flexibility and transactions costs issues. Such a contract will have some distribution of price fluctuation risk between buyer and seller. This distribution may not be Pareto optimal, because of different risk aversion or non-convexities in the firms' profit functions.^{4.2} But the firms can redistribute risk by supplementing the flexible long term contract with a forward contract for rigidly pre-specified price and quantity. The flexible long term contract then effectively determines only the price of deviations from the pre-specified quantity; the forward contract determines the average price.

The price per unit of the two contracts need not be the same. For example, the standard forward contract might be for 1000 tons each month at \$100 per ton, while the flexible long term contract has a price equal to a price index. If in a particular month the buyer decides it wants only 850 tons, it pays \$100,000 minus 150 times the current value of the price index. Production efficiency is unaffected by the standard forward contract since the marginal cost per ton to the buyer is the price index, not \$100, and the buyer chooses its quantity accordingly. Yet both buyer and seller are hedged.

This separation of production efficiency from risk sharing is highly desirable; it would not be achieved by offering a long term,

variable quantity contract such as is now standard for public utilities. For example suppose that a public utility offers to buy from independent suppliers at its current optimal spot price, and it also offers them a variable quantity long term contract arrangement at a price which changes less often than the spot price. Then the utility is essentially giving away an infinite sized commodity put option. The independent supplier will sell at the spot price or the price set by the long term contract, whichever is higher. Thus production efficiency is not achieved, and all risk of price fluctuations is born by the utility. In general, contracts of multiple duration may simultaneously be in force for transactions between two firms, but only one of them should have variable quantities. That contract will determine the marginal cost per unit, hence determine the level of joint production efficiency. The others should be standard forward contracts or at least pre-specify quantities, and be used solely to redistribute risk.

Third parties may also offer to take either side of standard forward contracts, as a purely financial transaction, if they can bear the risk at a lower cost than either of the original parties. In general risk shedding will not be costless and in thin markets it may have very high transactions costs. But to the extent risk is a concern of either the buyer or the seller, they can reduce their risk, at a price, without affecting the underlying production cost/transactions cost trade-off achieved via long term contracts with flexible quantities and perhaps prices.^{4.3}

2.5 A Formal Model of Optimally Sticky Prices

This section provides a formal model of the trade-off between transactions and production costs, as functions of the interval between price changes. The model is soluble in closed form, but is designed to give insights rather than quantitative results.^{5.1}

Later chapters will use these insights to help develop and solve analogous but detailed and market-specific models.

Problem Statement

Consider a welfare maximizing monopolist which sells a single homogeneous product. (With suitable redefinition of variables the same model will generalize to a profit maximizing monopolist and to the competitive case.) Both the marginal production cost and the demand for the product shift stochastically over time, for example due to changing input prices for the producer. These shifts lead to corresponding shifts in the "optimal" spot price $p^*(t)$ for the product at time t . This price is defined as the optimal price in the absence of transactions costs. In simple cases it is given by the intersection of instantaneous demand and marginal cost curves, as shown in Figure 2.5.1.^{5.2} Thus $p^*(t)$ traces out a path such as that in Figure 2.5.2.

The actual price set by the monopolist at time t is $p'(t)$. I will assume a simple selling arrangement, in which customers can demand as much as they want, and pay $p'(t)$ per unit purchased at t . Therefore whenever $p'(t) = p^*(t)$, joint production costs are not

Figure 2.5.1
Determination of p^* and L at time t

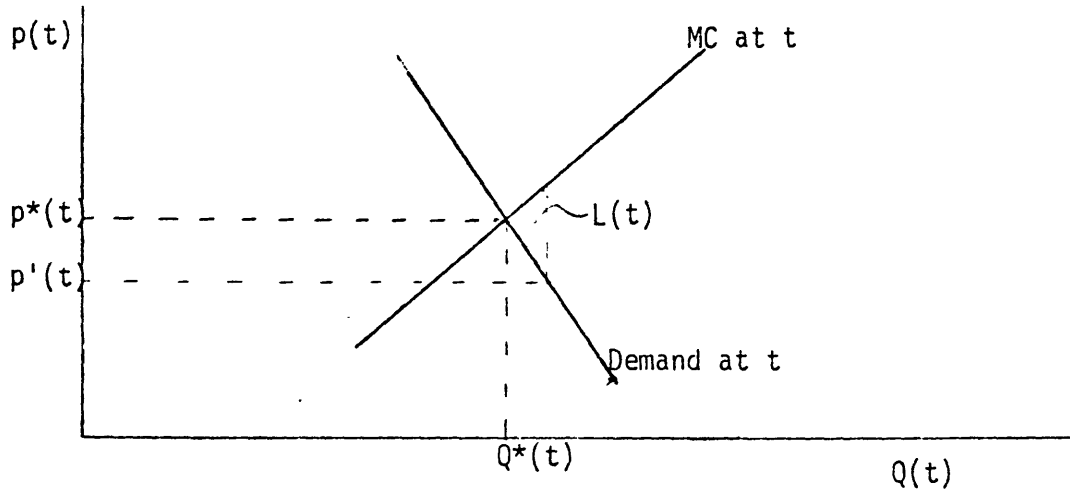
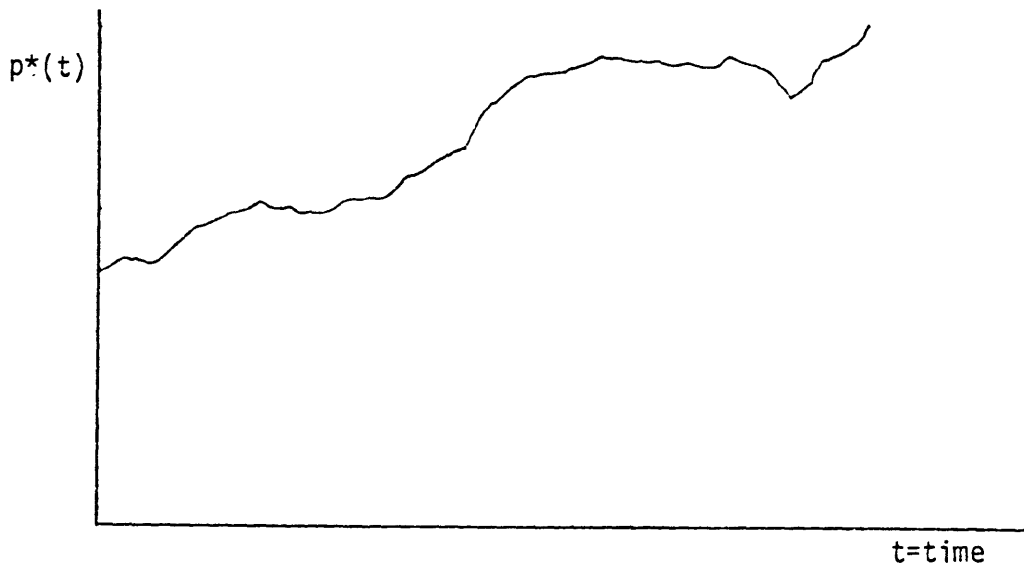


Figure 2.5.2
Sample path of $p^*(t)$ over time



minimized. Take as the objective maximizing producer's plus consumers' surplus. Call the loss of potentially available surplus $L(t)$. In Figure 2.5.1 it is indicated by the shaded triangle, for an example where $p'(t) < p^*(t)$.

Assume that $L(t)$ can be approximated by a quadratic function of the deviation between actual and optimal spot prices:

$$(2.5.1) \quad L(t) = L[p'(t), p^*(t)] = k [p'(t) - p^*(t)]^2$$

$$(2.5.2) \quad k = \frac{p^*(t) Q^*(t)}{2 [p^*(t)]^2} \times \epsilon^{\text{demand}}(t) \times \left[1 + \frac{\epsilon^{\text{demand}}(t)}{\epsilon^{\text{supply}}(t)} \right]$$

where

k = Coefficient on loss function

$Q^*(t)$ = Socially optimal demand at t (see Figure 2.5.1)

ϵ^{demand} = Demand elasticity at t (absolute value)

ϵ^{supply} = Supply elasticity at t .

(These formulas are exact only for linear demand and supply curves.)

Thus a given percentage price deviation causes more loss if the total market size p^*Q^* is large, if demand is very responsive to price, or if marginal costs of production are very sensitive to demand.

The producer can change $p'(t)$ at any time. If these changes were costless, optimal policy would be to have $p'(t)$ track the optimal spot price exactly, i.e. $p'(t) = p^*(t)$ at all t . However, each time $p'(t)$ is changed, transactions costs are incurred. Therefore it is optimal to have $p'(t)$ be somewhat "sticky", i.e. not track $p^*(t)$ perfectly.

I will model two ways to change $p'(t)$, with different

transactions costs. First, it can be recalculated and the new value immediately put into effect. In a bilateral negotiation situation, this corresponds to the expiration and renegotiation of a contract, and as discussed may have high transactions costs. Even in markets where potential opportunism is not a problem, such recalculations are not costless. Assume that the total transactions costs of n price recalculations are $n C_r$.

Second, $p'(t)$ can be changed according to a predetermined schedule. The schedule is updated at the time of each price recalculation. Examples of predetermined price changes are shift differentials in labor contracts, and time-of-day pricing of electricity, telephone calls, movies, and television ads. As I will show, these predetermined price changes are only useful for tracking predictable changes in $p^*(t)$. The transactions costs of predetermined changes are due to more complex rate schedules and billing procedures, i.e. to contract implementation costs. Assume that the total transactions costs of n predetermined price changes between each price recalculation is $n C_p$.^{5.3}

The producer now faces the following problem.^{5.4} At the time of each price recalculation, it must choose:

1. The time of the next price recalculation, i.e. the contract duration. Call this T_r .
2. The number and times of predetermined price changes between now and T_r . For simplicity I will examine only stochastic processes for $p^*(t)$ such that the optimal policy is to space

predetermined changes evenly, T_p ($< T_r$) apart.

3. The level of actual prices $p'(t)$ to take effect at each predetermined price change.

Figure 2.5.3 shows an example. Suppose one contract expired at $t=0$, and at that time values for T_p , T_r , $p'(0)$, $p'(T_p)$, and $p'(2T_p)$ were chosen as shown. Then $p^*(t)$ occurred as shown. As it turned out, $p'(t)$ deviated from $p^*(t)$. At time T_r , the choice process is repeated, using all new information about the $p^*(t)$ process.

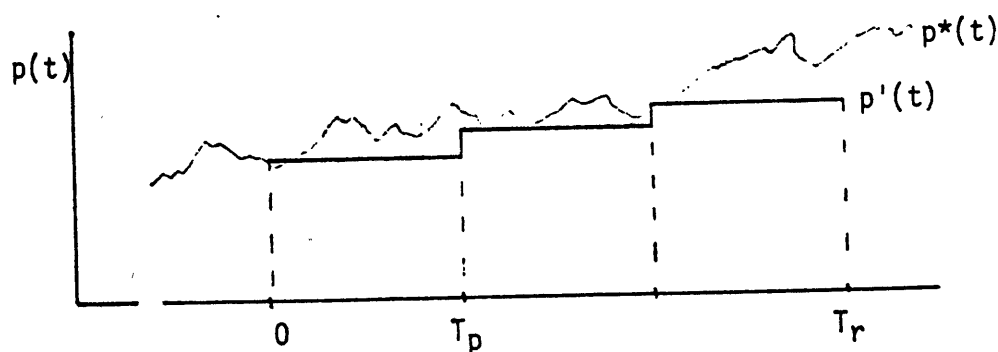


Figure 2.5.3

Hypothetical Pricing Policy and Realization of $p^*(t)$

The optimization problem at time 0 is thus to set T_r , T_p , and $p'(t)$ to minimize the average value of losses plus transactions costs per unit time:

$$(2.5.3) \quad \text{Min} \quad E \frac{1}{T_r} \left[\int_0^{T_r} L(t) dt + C_r + C_p (-1 + T_r/T_p) \right]$$

subject to:

$$(2.5.4) \quad T_p \leq T_r$$

where:^{5.5}

$$L(t) = k [p'(t) - p^*(t)]^2$$

k = sensitivity of losses to price deviations (eq 2.5.2)

$p^*(t)$ = "optimal" spot price at time t

$p'(t)$ = actual price at time t

T_r = interval between full spot change/recalculations

T_p = interval between predetermined price changes, with the new prices prespecified at the time of previous spot price recalculation.

C_r = Transactions costs of each spot price change/recalculation

C_p = Transactions costs of each predetermined price change

Models which look at the trade-off between transactions costs (without discussing what determines these costs) and production costs are not new. Barro [1972] and Gray [1978] covered the case when $p^*(t)$ follows a random walk with zero mean. Sheshinski and Weiss [1977] and Mussa [1981] assumed purely deterministic changes in $p^*(t)$. Rotemberg [1981] and Danziger [1981] allow both: random walks with a trend. All these authors have a similar objective function of minimizing production cost losses (or monopolist's profit; see below) plus transactions costs. Gray takes the production losses ($L(t)$) as quadratic in the log of output; the others take them as quadratic in output or price.

Allowable types of price adjustment vary by author. Barro uses an (S,s) inventory style rule.^{5,6} Sheshinski and Weiss, Gray, and Danzinger use the above "prespecified T_r ", rule, where T_r corresponds to contract duration. In all these cases, each price change has a fixed, discrete, cost. (Rotemberg allows continuous price adjustment with a transactions cost which is quadratic in the rate of price change. His results are therefore quite different from the others; prices change continuously, rather than in discrete jumps.) Gray and Danzinger also allow continuous, costless price adjustments if they are prespecified at the time of contract renewal. In Gray's case these are continuous indexation rules; in Danzinger's case all changes must be completely prespecified.

Thus Danzinger's model is closest to the model here. He has a slightly different demand function, supply function, and stochastic process for $p^*(t)$. The most important difference is that he assumes predetermined price changes are costless. In his model it is therefore always optimal to remove the predictable change in $p^*(t)$ via predetermined price changes. Also, his formulation does not lead to a closed form solution.

Alternate Interpretations

Although this model was stated for the case of a profit maximizing monopolist, it also applies in other cases, including the case of competition turning into bilateral monopoly, treated in Section 2.2. Figure 2.5.4 shows how variables should be redefined in these alternate cases.

Figure 2.5.4
Alternate Definitions of Variables

<u>Objective</u> <u>Function</u>	<u>$p^*(t)$</u> <u>spot price</u>	<u>$L(t)$</u> <u>loss</u>	<u>C_p, C_r</u> <u>transact. costs</u>
Social welfare max	$p^*(t)=MC(t)$	Loss of CS+PS	Implementation cost
Profit max monopolist	$MR(Q(p^*))$ $= MC(t)$	Loss of PS + Monopoly profit	Producer implement. cost + loss due to customer fear ^{5.7}
Mutual idiosyncrat. investments	$p^*(t)=MC(t)$	Loss of CS+PS	Implement. costs + Expected value of mutual opportunism at end of contract

where:

- $p^*(t)$ = Optimal spot price for incremental purchases
- MC = Marginal cost of production
- MR = Marginal revenue of monopolist
- $Q(p)$ = Customer demand at price p
- CS = Consumers' surplus
- PS = Producer's surplus

Solving the Model

This problem is easily simplified by solving for the optimal $p'(t)$ as a function of T_p , then solving for T_p and T_r . Losses can be decomposed into a forecast error component and a time aggregation error component. Define $p^m(t) = E p^*(t)$. Then:

$$\begin{aligned}
 (2.5.5) \quad L(t) &= k [p^*(t) - p'(t)]^2 \\
 E \int L(t) dt &= k \int E [p^*(t) - p'(t)]^2 dt \\
 &= k \int \text{Var} [p^*(t)] dt + k \int [p^m(t) - p'(t)]^2 dt \\
 &= \text{Loss due to inability to forecast } p^*(t) \\
 &\quad + \text{Losses due to not setting } p'(t) = p^m(t) \\
 &\quad \text{exactly.}
 \end{aligned}$$

Here the variance and expectations are conditional on all information known at the time $p'(t)$ is set. The first term is zero if there is no forecast error in the forecast of $p^*(t)$. The second term is zero if there is no time aggregation error of prices, i.e. if $p'(t) = p^m(t)$ for all $t \in [0, T_r]$.

Plugging this into the maximization problem (2.5.3) gives a reduced problem:

$$\begin{aligned}
 (2.5.6) \quad \text{Min } E_0 \quad &k \int \text{Var } p^*(s) ds + (C_r - C_p) + \int [C_p + k [p^m(s) - p'(s)]^2 ds] \\
 &+ [C_p + k \int [p^m(s) - p'(s)]^2 ds] + \dots + [C_p + k \int [p^m(s) - p'(s)]^2 ds]
 \end{aligned}$$

subject to (2.5.4). Here T_r/T_p is the number of predetermined price changes which will be needed before the next price recalculation, including the price change at time 0.

Notice that this is two almost independent subproblems, one involving T_r and the other T_p and the various $p'(t)$. They are connected only by constraint (2.5.4). Finding T_r requires considering only the forecast error, while finding optimal T_p involves only time aggregation error. Time aggregation error accumulates only until the next predetermined price change; hence the different limits on the integrals.

Now find the optimal predetermined prices as a function of T_p . Since price must be held constant over the interval $[t, t+T_p)$ and $p'(t)$ must be selected at time 0, then solving the relevant portion of (2.5.6) gives the first order conditions:

$$0 = \frac{\partial}{\partial p'} [p^m(s) - p'(t)]^2 ds$$

$$(2.5.7) \quad \text{Optimal } p'(t) = \int_t^{t+T_p} p^m(s) ds$$

= Average expected price $p^*(s)$ during the interval $[t, t+T_p)$.

Result: Prespecified prices should be set to expected optimal spot price, averaged over the interval the prespecified price will be in effect.^{5.8}

At this stage it is convenient to be more specific about the stochastic process $p^*(t)$. For analytic purposes, I will use a Wiener process:

(2.5.8) $dp^* = \mu dt + \sigma dz$ where z is Brownian motion. Thus

$$(2.5.9) \quad p^m(s) = E_t p^*(s) = p^*(t) + \mu(s-t) \text{ for } s > t.$$

A Wiener process is an adequate approximation to any continuous stochastic process over a short enough interval, but its primary justification here is that it leads to analytic solutions.

Using (2.5.9) in (2.5.7) gives the optimal predetermined prices set at time 0 as:

$$(2.5.10) \quad p'(0) = p^*(0) + \mu T_p / 2 = p^m(0) + \mu T_p / 2$$

$$p'(n T_p) = p^*(0) + \mu T_p (2n+1)/2 = p^m(n T_p) + \mu T_p / 2$$

Having found the optimal predetermined price as a function of T_p , we can now solve for T_p . Unfortunately, the character of the solution depends on whether the constraint (2.5.4) binds. First consider the case where it does not, i.e. it is optimal to use at least one predetermined price change before the next price recalculation. Differentiating (2.5.6) gives the first order condition for T_p , the optimal interval between predetermined price changes:

$$(2.5.11) \quad \frac{C_p}{T_p} = k \sum_{n=1}^{T_r/T_p} [p^m(n T_p) - p'(n T_p)]^2 - \frac{1}{T_p} \int_{(n-1)T_p}^{n T_p} [p^m(s) - p'(s)]^2 ds$$

Average = transactions cost per unit time	Time aggregation error losses at time of next predetermined price change	-	Average time aggreg. error losses over interval between price changes
--	---	---	--

Specifically, if we plug in the Wiener process for $p^m(t)$ and the resulting optimal $p'(t)$ from (2.5.10) we see that each element of the summation on the right hand side of (2.5.11) is equal. After some algebra we get for optimal T_p :

$$(2.5.12) \quad T_p = \left(\frac{6C_p}{k\mu^2} \right)^{1/3}$$

Result: The optimal interval between predetermined price changes depends on the transactions costs of such changes compared with the predictable rate of change of optimal spot prices. The uncertainty of optimal spot prices and the interval between price recalculations are irrelevant, as long as it is optimal to have predetermined price changes in between spot recalculations.

Optimal Contract Duration

Similarly, we can find the optimal interval between price recalculations/contract duration, T_r . Differentiating (2.5.6) gives the first order conditions:

$$(2.5.13) \quad \frac{C_r - C_p}{T_r} = k \text{ Var } p^*(T_r) - \frac{k}{T_r} \int_0^{T_r} \text{Var } p^*(s) ds$$

Average cost per unit time of a recalculation	=	Losses due to forecast error at T_r	-	Average forecast error over the interval $[0, T_r)$
--	---	---	---	---

For the Wiener process we get:

$$(2.5.14) \quad T_r = \left[\frac{2 (C_r - C_p)}{k\sigma^2} \right]^{1/2}$$

Lemma: If predetermined price changes are used between price recalculations, then the optimal interval between price recalculations depends only on the forecast variance of optimal spot prices, but not on predictable changes in them.

Now consider the case when (2.5.4) does bind, i.e solving (2.5.11) and (2.5.13) would violate the constraint that $T_r \leq T_p$.

Then (2.5.10) is solved by finding T_r such that:

$$(2.5.15) \quad C_r/T_r = L(T_r) - \frac{1}{T_r} \int_0^{T_r} L(s) ds .$$

For the Wiener process this has the cubic "solution":

$$(2.5.16) \quad T_r = T_p = \frac{2 C_r}{k [\sigma^2 + (T_r \mu^2)/3]}$$

In this case forecast variance and predictable changes in optimal spot prices both affect optimal contract duration/interval between price changes.

Optimal Price Changing

We are now in a position to step back and look at optimal patterns of price change.

Result: If optimal spot prices change continuously, and there is a positive cost for each time actual prices are changed, then optimal actual prices will be "sticky", and will therefore deviate from underlying optimal spot prices.

Proof: The Wiener process case is easily proved from (2.5.15). The more general case is proved by noting that continuous changes in actual price would have infinitely high transactions costs.

Comparing (2.5.12) and (2.5.14), we can see under what conditions it will be optimal to use only price recalculations, i.e. contracts without intermediate price adjustments. It turns out that optimal behavior is a discontinuous function of market conditions.

Result: If C_r is close to C_p , i.e. predetermined price changes are almost as expensive as spot changes, or if forecast variance dominates predictable changes in p^* , then predetermined price changes should not be used, only spot changes. If μ , the predictable component of $\partial p^*(t)/\partial t$, increases, this will lead to a drop in the optimal interval T_r between price recalculations. Ultimately it will cause a discrete switch to a mix of predetermined and spot price changes. Thereafter higher values of μ will only affect the interval between predetermined price changes. ^{5.9}

Figure 2.5.5 summarizes the sensitivity of optimal contract duration and price stickiness to characteristics of the market and its transactions costs. For example, the larger a market, the more frequently prices should be adjusted, all else equal. Conversely, the larger transactions costs the less often prices should be adjusted.

Figure 2.5.5
Sensitivity Analysis

<u>Parameter</u>	<u>Effect on optimal interval between price changes</u>		
	T_r	T_p	$T_p = T_r^*$
σ =Stochastic change rate	+	0	+
μ =Predictable change rate	0	+	+
C_r =Recalculation cost	+	0	+
C_p =Predetermined adjustment cost	-	+	0
Q^*p^* =Market size	-	-	-
ϵ_{demand}	-	-	-
ϵ_{supply}	+	+	+

*This column applies if it is optimal to use only price recalculations. (Equation 2.5.16)

Indexed Prices

Previous analysis assumed that prices were either not adjusted at all during the course of a long term contract, or were adjusted to predetermined levels. As discussed in Section 2.2, sometimes it will be desirable to tie price $p'(t)$ to the level of a relevant index. This will be optimal if the costs of measuring the index are low, the index is highly correlated with $p^*(t)$, and the index is essentially beyond the control of either party, so that it does not create opportunistic behavior.

If a price index is used, the optimal interval between indexed adjustments will depend on the transactions cost per adjustment compared with the forecast variance of the indexed portion of $p^*(t)$. A more accurate index, or a lower transactions cost per indexed price adjustment leads to optimally more frequent adjustments.

The use of a good price index will increase the optimal contract duration T_p , since it will on average decrease the value of $p^*(t) - p'(t)$ for any t . In the limit, as the index approaches a perfect predictor of $p^*(t)$, optimal contract duration approaches infinity. Thus Congress did well to tie a variety of transfer payments to inflation indices; but it would not be optimal to avoid legislative adjustments to them indefinitely.

2.6 Some Implications for Pricing in Different Markets

Optimal price changing policies, then, are a function of market characteristics. In some markets optimal spot prices calculated by standard static models change slowly and predictably, and transactions costs are low, so that optimal pricing policy is to track the optimal spot price closely. For such markets conventional models which ignore transactions costs are reasonably accurate. Other markets are at the opposite extreme, and we predict that in such markets prices will be quite "sticky" compared with the underlying optimal spot prices. Figure 2.6.1 summarizes the characteristics of markets which should have sticky prices.

What price adjustment methods will be used? At least three are available: price adjustments based on an exogenous index, spot recalculations at predetermined intervals, and changes of predetermined magnitude at variable intervals (the (S,s) method). Each method has different transactions costs and different ability to track optimal spot prices in different markets. Often multiple methods will be used. [Stigler and Kindahl p. 35]^{6.1}

The choice of methods depends on transactions costs and on how closely each method can track p^* . Transactions costs are dependent on the level of potential opportunism. Indexed price adjustments are least susceptible to opportunistic manipulation, subject to caveats discussed earlier, since once the indexation formula is set, no further discretion is available to either party to affect the price. Conversely, (S,s) rules have high transactions costs when opportunism

Figure 2.6.1

Market Characteristics Leading to "Sticky" Prices

- o Rapidly changing optimal spot prices, which are caused by any of the following:
 1. Large and rapid swings in instantaneous supply and demand curves.
 - Heavy dependence on exogenous random variables (weather, fuel prices, etc.)
 - No good substitutes, or;
 2. Product storage by producers or arbitragers is expensive.
 3. It is expensive for users to shift their demands in time.
 4. Short run marginal costs of production are steeply increasing.
 5. Production capacity is often fully used. (Capital intensive industries.)
- o Transactions costs of price changes in this market are high, which will happen if any of the following hold:
 1. Idiosyncratic product (small number of buyers or sellers).
 2. High costs of informing customers about price changes.
 - Large number of small customers.
 - No natural communications channel such as trade journals or visible signs at point of sale.
 3. Exogenous price indices inaccurate, expensive to monitor.

is a problem, since the decision to change prices requires continual negotiation about the current level of p^* . Full recalculations at predetermined intervals cause an intermediate level of transactions costs, since p^* must be agreed on only at intervals.

How closely do the three methods track p^* ? Indexation tracks closely to the extent that p^* is a predictable function of exogenous variables. (S,s) rules ensure that p^*-p' never becomes extremely large, yet do not require price changes when p^*-p' is small. The tracking ability of recalculations at predetermined intervals depends on the interval length.

Application to Electricity

What are the implications of the above analysis for how electricity, in particular, should be priced? Here I develop qualitative hypotheses; the numerical welfare analysis will be covered in subsequent chapters.

Electricity prices have changed faster over the last decade, in response to faster changes in p^* . But there is no reason to think that current price adjustment procedures are optimal, since they evolved in response to political pressures instead of being deliberately designed to make the optimal trade-off between transactions costs and production costs. The economic literature on time-of-use pricing has been a step toward systematic design of price-changing policies, but has made arbitrary assumptions. First is the assumption that only predetermined price changes are available.

This ignores the stochastic variation in $p^*(t)$, which for electricity is quite large even from one day to the next (see Section 5.1).

Second, there has been little effort to find the optimal interval between predetermined price changes. Typical analyses compare flat prices with one time-of-use pattern instead of comparing a spectrum of time-of-use patterns. This point is made by Koenker [1979, p. 180, 187].

What are the implications of this chapter for optimal pricing patterns in different electricity markets? The implications of different stochastic processes for $p^*(t)$ are apparent from Figure 2.6.1 and Section 2.5. For example, utilities where oil is always the marginal fuel will have little variation in $p^*(t)$ and therefore (all else equal) longer optimal intervals between price changes than will utilities which shift daily from oil to coal on the margin. Utilities where demand and supply follow rigid predictable patterns should rely more on predetermined price changes, and have less frequent price recalculations.

The implications of transactions costs are more complex. Implementation transactions costs (metering, billing, etc.) are roughly proportional to the number of customers affected, and are an increasing function of the number of predetermined price changes. Considering the impact of the ratio C_p/p^*Q^* in section 2.5, this says that large customers should have more frequent predetermined price changes than equally elastic small customers. This conclusion is not controversial; most states which have implemented time-of-use

rates have started with large customers. [Malko and Faruqi, 1980] Yet when time-of-use rates have been implemented for several customer classes in the same utility territory, each rate schedule usually has the same pricing periods. This appears non-optimal.^{6.2} In particular, many industrial customers are already on detailed (hourly or more often) meters. The cost of implementing predetermined hourly price changes for such customers would be quite small.

The interval between price recalculations affects both contracting transactions costs (which are affected by opportunism) and implementation transactions costs. When opportunism is an issue, it is therefore hard to rigorously determine the optimal interval between recalculations. However there are special cases in which a utility would have little or no incentive to opportunistically manipulate prices, implying that in these situations more frequent recalculations are optimal. Buybacks of energy from independent producers are one such case. A utility will want to minimize the price of such buybacks; therefore if the price is properly calculated (see Chapter 3) the utility has no way to manipulate the price in its favor. Hence a buyback rate should specify more frequent recalculations than a standard sell rate, for an energy sale of the same magnitude.

Finally, it is possible to design rates so that a utility has no additional incentive to manipulate prices beyond what is built into the current system of rate hearings and fuel adjustment charges. For such rates the interval between price recalculations should be determined solely by the trade-off between production costs and

implementation costs. (This will be formalized in Chapter 3.) This suggests much closer tracking of the optimal spot price for large and responsive customers than for small or unresponsive ones. In special cases such rate designs may lead to substantial efficiency improvements, as will be shown in Chapter 5.

FOOTNOTES TO CHAPTER 2

0.1 My thanks for useful comments on this chapter to Paul Joskow, Bruce Kogut, Nalin Kulatilaka, Julio Rotemberg, Richard Schmalensee, and Fred Schuppe. Since I have not always followed their advice, I alone am responsible for any errors.

0.2 Many of these contracts had adjustment clauses, so contract duration is not equal to the interval between price changes.

0.3 An alternate explanation for "sticky" or "jumpy" prices is that marginal costs or other factors affecting the optimal price in a static model also change in a jumpy fashion, and prices immediately adjust fully. However the costs of most products are functions of the prices of multiple inputs, each of which changes irregularly. Hence this explanation predicts price changes at intervals from every few minutes to (at most) days for most products.

0.4 Price changing patterns have many dimensions, not just the one shown in Figure 2.0.2. The most important dimensions will be discussed in section 2.5.

1.1 Roughly speaking, the optimal spot price is the intersection of the supply and demand curves in a static model. This is not exact since conventional static equilibrium models do not capture the joint nature of production and consumption at different times. The next chapter shows how to calculate precisely the optimal spot price (defined when price adjustments are costless) even if there are intertemporal production or use relationships.

1.2 Barro [1977, p. 311] makes this point about labor markets. Fischer's response [1977] points out that such "Pareto superior" contracts are, in fact, not seen in labor markets. As subsequent discussion will show, the same remarks could have been made about other markets.

1.3 The formula for the optimal spot price at each moment will depend on the motives of and constraints on the price setter. It may be the outcome of a monopolistic, weak bilateral monopoly, regulated monopolistic, welfare maximizing, or even oligopolistic model, each of which can lead to different prices. The relevant price is the price of the last unit purchased. A complexity is that if the marginal cost of production is rapidly upward sloping relative to the size of the buyer, the optimal price may depend on the buyer's own demand level. Then the buyer should be given a price schedule, not a single price. I will ignore this issue for now; think of the optimal spot price as a one dimensional parameterization of a price schedule.

1.4 Even at time t there may still be uncertainty about the "ideal" price. For example Lodish [1981] models the effect of uncertainty about future demand on the optimal price now in sales of advertising spots. Thus p^* may be the result of an optimizing procedure which yields a single price despite uncertainty. All results then go through. My thanks to N. Kulatilaka for pointing out this issue.

2.1 The opportunism problem is not relevant to social welfare maximizing monopoly models, since by assumption the monopolist has no incentive to take advantage of others' investments once they are made.

2.2 These mechanisms are adapted from Williamson [1979]. Obviously I am responsible for any errors I have introduced.

2.3 Williamson feels that quantity adjustment with prices fixed gives little incentive for opportunistic behavior (p. 251). But it seems to me that such adjustments are only safe in one direction: when the buyer wants to reduce the quantity purchased at the fixed price (and can show that he is not turning to another supplier). Unilaterally imposed cutbacks by the seller could simply be attempts to force a higher price. Unilaterally imposed increases by the seller, at the same price, could be "dumping" of a product. Allowing the buyer to increase the quantity at a fixed price gives it a potentially extremely valuable open-ended option, unless arbitrage of any kind can be ruled out.

2.4 If average cost pricing is used for all units sold this is a deliberate deviation from optimal spot prices. However it is possible to come close to optimal prices at the margin, while still limiting profits. See Chapter 3, and the next footnote.

2.5 For example, for mechanism 4, it is the seller's total profit which is relevant, rather than its profits from sales to each individual purchaser. Buyers can remove the incentives for opportunism by monitoring and controlling total profits. Hall [1981] proposes an interesting way of rebating any resulting excess profits. He does not discuss the need for a multilateral contract to implement his proposal.

3.1 Customers will favor a single buyer's agent for two reasons. First, there are scale economies in negotiating and monitoring a contract. Second, using a single agent gives customers countervailing monopoly power.

3.2 This point is made in Baron and Taggart [1980]. "The procedures that regulatory authorities utilize to set prices, if not

based solely on exogenous factors and on estimated or ideal factor inputs, thus can create an incentive for a firm to take strategic actions to further its own interests...through prices that diverge from the regulatory ideal or technical inefficiency or both." [p. 41]

3.3 At first glance, prices based on long run marginal costs might seem to be an exception; but they are not. Suppose prices are based on long run marginal costs, presumably with time of day changes. There is too much uncertainty in the world to leave the prices alone indefinitely. As prices are revised, the only way to avoid possible opportunistic behavior by the utility is to set prices based on an "ideal" utility, with an "optimal" capital stock, set of maintenance and fuel purchase procedures, and so on. This would require the price-setters to know the utility's current and potential production function, or to use a real outside utility for reference. Any real outside utility will have different load characteristics and different site characteristics, and hence not be comparable. Knowing the production function is also unrealistic. Therefore the prices would have to be revised looking at actual and potential behavior by the utility, and again the possibility of opportunism arises.

3.4 Vickrey [1971] proposes a multiplier on this fund to give the utility incentives to build more plants. The properties of his multiplier, however, may not be desirable.

3.5 If the utility has an ex post optimal capital stock (given the actual value of fuel prices, nuclear licensing, etc.) and if regulators use the proper discounting formulas, and if there are no scale economies in constructing plants or transmission lines the fund would have an expected revenue of zero each year if full spot pricing is used. That is, quantity times full spot price, integrated over space and time, equals total costs. See Chapter 3. In fact, of course, only luck will give a utility a capital stock which is ex post optimal. However unless capital stock is grievously incorrect, the absolute value of the revolving fund should be small.

4.1 See for example Holthausen [1979]. Beware his conclusion that firm output decisions depend only on forward prices if forward contracts are available; this does not generalize beyond the two period case.

4.2 Bankruptcy costs are such a non-convexity. This is not to say that situations with non-convexities or differential risk aversion are important or common for most products. Most risk of price fluctuation will be unsystematic risk, in the sense of the Capital Asset Pricing Model, and therefore irrelevant to publicly owned firms which are not in any danger of insolvency. But see the

next footnote.

4.3 This is too sanguine a view for labor markets. In these markets sellers are definitely risk averse, and standard forward contracts are not possible because of moral hazard. Hence risk may have a significant effect on the form of labor contracts. But transactions costs also affect the arrangements for dealing with risk in labor markets; see Wachter and Williamson [1979] for a discussion. Since the major concern of this thesis is public utilities, I will not discuss labor markets further.

5.1 Later chapters provide a much more detailed model for one market (electricity) and solve it using numerical search.

5.2 The conditions under which Figure 2.5.1 is sufficient to determine $p^*(t)$ are that demand and supply in successive periods each have zero cross-price elasticities. This condition does not hold in real markets; Chapter 3 will use a more general model.

5.3 Since each price recalculation involves one change in the current price, it must be that $C_p \leq C_r$.

5.4 A third price adjustment method is to let the level of the next price adjustment be fixed, but its time of occurrence not fixed. This is like an (S,s) inventory adjustment rule--see Barro [1972]. This method substitutes for price recalculations of the "fixed interval" type. Which is optimal depends on their comparative transactions costs and tracking ability. When opportunism is a problem, the (S,s) rule may have high transactions costs. And for some products, such as electricity, implementation costs depend mainly on how often prices can be changed, not on how often they actually are changed. Hence the (S,s) policy for electricity will often be dominated by the "fixed interval" policy.

Adjustments made using the (S,s) methods will have similar properties to those made using the "fixed T_r " method. In one case the amount of each price adjustment is variable but the time is fixed; in the other the converse. Statements in the text about "optimal duration" still hold on average for the (S,s) method, instead of for certain as in the "fixed interval" method. See also the discussion in Section 2.6.

5.5 I ignore here the restriction that T_r should be an integral multiple of T_p , or if it is not integral then the formulas here are not exact. Thus in practice selecting T_p will involve a choice among discrete levels, rather than the continuous spectrum assumed here. This refinement would remove the ability to use calculus and adds little here. In chapter 5 I will observe the restriction, however.

5.6 See footnote 5.4.

5.7 A monopolist may offer a long term contract, to induce the buyer to make product-specific investments and thereby increase its demand.

5.8 Section 3.3 will show that this is approximately but not exactly correct with more general loss functions. It is a familiar result in the time of day pricing literature; see the literature review in section 3.4.

5.9 It would be interesting to test these results on historical data about contract duration as a function of inflation.

6.1 For example, union labor contracts often combine full recalculation/renegotiation at three year intervals, with annual indexed adjustments, and apparently with the possibility of unscheduled renegotiation in mid contract if the optimal spot price falls too far below the contractually set wage. [Business Week 1981, especially p. 89]

6.2 Unless it can be explained by perceived equity issues or other political constraints. Yet variations in average cost per kwh between different rate classes typically have income redistribution implications which dwarf the effect of more finely tuned pricing periods for certain customers. Hence political constraints do not appear to be a likely explanation.

CHAPTER 3

OPTIMAL ELECTRICITY PRICES

Exactly how should a public utility set prices? Chapter 2's analysis divides this into two stages. First, it invoked the existence of underlying "full spot" prices, which vary across space, time, and states of nature. Second, it considered the implications of transactions costs caused by price changes. The existence of these costs means that both socially optimal and competitive prices will not "track" the full spot prices exactly. Instead they will be somewhat aggregated across space, time, and states of nature. The optimal amount of aggregation depends on the trade-off between transactions costs and misallocation caused by deviation from the full spot prices. Different amounts of aggregation give different rules for setting actual prices, or "rates."¹ The purpose of this chapter is to develop both stages of this analysis rigorously for public utilities provided through a transmission network. First, the chapter develops equations for socially optimal full spot prices. Various complexities are introduced gradually. Optimal dispatching and investment rules are derived easily using full spot prices. Second it shows how, for a given level of aggregation along the time state-of-nature or spatial dimensions, actual prices should be set. Third, it shows how to determine the socially optimal level of price aggregation for a particular customer. Finally, it considers how to control opportunistic behavior by a monopolistic public utility.

Throughout this chapter, I use generalized neoclassical production functions and related concepts. This simplifies notation and derivations, but it begs a crucial issue: How can customers respond to prices which change rapidly and stochastically? If they cannot respond rapidly, or do not find it cost-effective to do so, then full spot pricing is pointless. Chapter 4 will address this by "opening up" firms' production functions to model real time response to stochastically changing prices, using a linear programming formulation.

This chapter will use the terminology of electricity. But the same concepts apply to any public utility, especially if it is centrally produced, then transported through a network. Natural gas fits the model almost exactly; municipal water supply less exactly.^{0.1} Most of the analysis applies equally to a monopolistic social welfare maximizing public utility and to a decentralized, more-or-less competitive industry. Therefore it can be considered a possible prescription for how to deregulate electricity generation, and a model of pricing in many commodity markets. I will concentrate on the social welfare maximizing case, and discuss decentralization briefly in Section 3.2.

The analyses of this chapter are closely related to the extensive literature on time-of-use pricing, as will be discussed in Section 3.4. That literature has shown that optimal prices across time are variable. However, it has generally ignored pricing as a function of spatial and state-of-nature (uncertainty) dimensions. Therefore this chapter provides solutions for several important

problems which are necessarily given unsatisfactory treatment in the existing literature. These include:

- o Rates for irregular purchases.
- o "Wheeling" rates, i.e., charges for the use of transmission capacity.
- o Buyback rates for energy purchased by a central utility from its customers, when supplies by customers are stochastic.

The solutions derived here appear quite unconventional and unfamiliar to utilities and public utility commissions. Therefore even if they "maximize social welfare," they may not be adopted. This suggests that the standard economic welfare maximizing formulation used here may be inadequate, because it overlooks legal and behavioral constraints. Nonetheless it is an important first step.

This chapter models the following elements:

- o Individual electricity users, with spatially distinct, stochastic, price-sensitive demands. Demands may be interdependent over time.
- o Individual generators, with spatially distinct, stochastic production levels. Most of these are under the control of a cost-minimizing central utility. Some may be controlled by individual profit-maximizing firms. Production may be interdependent over time.
- o Transmission and distribution of electricity, including line losses, line limits, and stochastic line outages.
- o Rationing of users and generators.

- o Capital investment in transmission, generation, and customer end use equipment.
- o Reactive energy will not be modeled here. See Caramanis et al. (1982) for this.

The perspective adopted for most of the chapter is the standard one of a global social welfare maximizing utility. This utility can set prices with any level of aggregation. It can ration participants (users and independent generators) in real time, but it must do so without complete information about the current state of each participant. It builds and operates the transmission and distribution (T and D) system and most generators. It has no direct control over investments by participants, but of course their investments will be affected by the prices they anticipate, and the central utility can influence those anticipated prices since it controls actual prices.

The organization of the chapter is shown in Figure 3.1. Section 3.1 presents and solves a simplified model of optimal spot prices, investment, and generation behavior. The Appendix to this Chapter is a more complex model which covers rationing, T and D system limits, and independent generators. Section 3.2 discusses results from the Appendix, without proving them. It discusses the concept of a partially decentralized utility system, with some generating units centrally owned, and others operated independently under spot prices. Transactions costs are ignored in Sections 3.1 and 3.2.

Section 3.3 derives optimal rates given that the prices must be aggregated over time or states of nature. That is, it shows how to

derive predetermined prices as functions of anticipated spot prices. The actual derivations are tedious and special cases have been developed by others, so this material is only sketched.

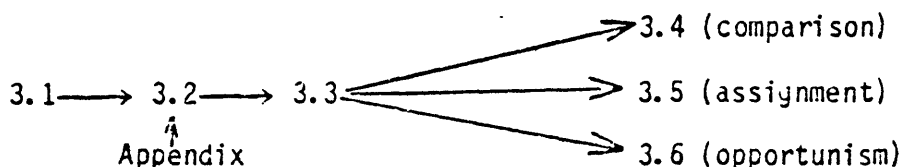
From here three independent branches are pursued. Section 3.4 relates spot pricing to the extensive literature on time-of-use pricing. It shows that optimal prices and investments under time-of-use pricing can be derived neatly using the concept of spot prices.

Section 3.5 considers transactions costs and how customers should be assigned to various rates other than full spot pricing. Unfortunately, neither mandatory nor voluntary assignment will always give the socially optimal assignment pattern.

Finally, Section 3.6 drops the assumption that the central utility is a welfare-maximizer. Instead, it discusses profit maximization under regulation. If prices are regulated and based on marginal costs, a profit-maximizing utility will have incentives to raise its costs, i.e., to produce inefficiently. This is true under spot pricing or time-of-use marginal cost-based pricing, and is different than the Averch-Johnson incentive problem. Section 3.6 discusses a class of regulatory mechanisms for eliminating this misincentive, albeit at a cost.

Figure 3.1

Outline of Chapter 3



Setting rates involves a conflict among multiple objectives.

These include:

- o Encouraging optimal behavior by customers and independent generators.
 - Short run (operating) behavior
 - Long run (investment) behavior
- o Reducing transactions costs
- o Controlling cross subsidies among participants
- o Encouraging optimal assignment of participants to rates
- o Controlling utility profits and opportunism.

Sections 3.1 and 3.2 derive rates which satisfy only the first objective. Subsequent sections consider how rates should be modified to encourage other objectives.

3.1 A Simple Model

This section presents and solves a simplified version of the optimal pricing problem for a central, welfare maximizing, utility. Assuming no transactions costs, full spot prices are optimal and are derived. The simple model is adequate to understand most of the key issues, although it is too simple for use by an actual utility. Section 3.2 will discuss the various complications of a full model, and present the results of the full model solved in an appendix.

Model Formulation

A utility system is composed of centrally owned and controlled generating plants, independent customers, and the transmission and distribution (T and D) system which links them. The utility must choose:

- o The output of each of its generating units.
- o The price to each customer.
- o Investments in future generating plants and the T and D system.

The utility must make these decisions to meet the following constraints and objectives:

- o Total generation must equal line losses plus total demand at each moment.
- o No generating unit can have an output higher than its available capacity.

- o Demands and unit availability vary stochastically.
- o Optimal dispatching: the utility sets output from each available unit to minimize short run operating costs, subject to constraints.
- o Optimal pricing: the utility sets prices to each customer to maximize total social welfare subject to all constraints.
- o Optimal investment: the utility plans investments to maximize social welfare in the long run.

For this section I will assume that all customers are under full spot pricing. That is, the utility can change their prices as often as it changes generator output settings, every few minutes. For now consider that the utility's capital stock is fixed; investment will be discussed later.

I now give simple mathematical models of each element of the utility system. These lead to a constrained optimization problem whose solution gives optimal full spot prices for each customer, and optimal output for each generating unit. Most of the material in Sections 3.1 and 3.2 was developed jointly with M. Caramanis and F. Schweppe, and is presented in Bohn et al. [1981] and Caramanis et al. [1982]. The model in this section is simplified. Figure 3.2.1 shows the differences between it and the more complete model in the Appendix.

Generation: The utility owns J generating units, each with maximum output K_j , deterministic marginal generating cost λ_j , and availability $\tilde{a}_j(t)$ during period t . For convenience units are numbered in order of operating cost, i.e. $\lambda_1 \leq \lambda_2 \dots \leq \lambda_j$.

Costs and unit availability in each period are independent of all other periods. Unit availability is an exogenous stochastic random variable between 0 and 1 which places a limit on generator output. Let $Y_j(t)$ be output from unit j at t , a decision variable. Then it is constrained to:

$$(3.1.1) \quad 0 \leq Y_j(t) \leq K_j \tilde{a}_j(t) \quad \forall j$$

Demand: Individual customers act independently, in response to time-of-day, weather, the price of electricity, the price of other inputs, and so on. In reality a customer may be either a firm, a household, or a neighboring utility. I will model all customers as price-taking expected profit-maximizing firms.^{1.1} Let F_i be the value added function for customer i 's use of electricity. It depends on the customer's electricity use $D_i(t)$ and on the random "weather" variable $\tilde{w}(t)$ which reflects exogenous economic and weather variables. Thus $F_i = F_i(D_i(t)/\tilde{w}(t))$, and if faced with price $p_i(t)$ at time t , the customer will choose $D_i(t)$ to maximize its consumer's surplus:

$$(3.1.2a) \quad \text{Consumer's surplus for } i = F_i(D_i(t)/\tilde{w}(t)) - p_i(t) D_i(t)$$

$$(3.1.2b) \quad \text{Hence} \quad \frac{\partial F_i(D_i(t)/\tilde{w}(t))}{\partial D_i(t)} = p_i(t) \quad \text{because of customer}$$

profit maximization. Since $\tilde{w}(t)$ is experienced by all customers, their demands will be correlated. Note that $D_i(t)$ may sometimes be negative, i.e. a "customer" may be a net producer of electricity.^{1.2}

Transmission and Distribution: The T and D system influences electricity losses and the variation in optimal prices across space. For convenience I will assume that all generators are centrally located, on a single infinite bus with no losses. Customers are dispersed along a "lossy" T and D system of arbitrary structure. Let $L(t)$ be the total losses throughout the transmission system. Then $L(t)$ is approximately quadratic in the demands [Elgerd, 1971, p. 297]:

$$(3.1.3) \quad L(t) = \underline{D}'(t) \underline{B} \underline{D}(t)$$

where

$$\underline{D}(t) = \langle D_1(t), \dots, D_I(t) \rangle$$

= Vector of demands.

\underline{B} = loss matrix, which depends on the location of customers and the shape and strength of the T and D system. It is not a diagonal matrix.^{1.3}

The utility faces an energy balance constraint:

$$(3.1.4) \quad G(t) = \sum_i D_i[t, \tilde{w}(t), p_i(t)] + L(t)$$

where

$$(3.1.5) \quad G(t) = \sum_j Y_j(t) = \text{Total generation at time } t$$

Violating this constraint significantly will cause an almost immediate uncontrolled blackout.^{1.4}

Optimization Problem: In the short run, the utility has a fixed capital stock K_1, \dots, K_j . Social welfare corresponds to customers' value added minus the utility's costs, which is the same as conventional consumers' plus producers' surplus.^{1.5}

The short term expected welfare maximization problem at period t can then be written as a Lagrange multiplier problem:

$$(3.1.6) \quad \text{Max } E_t \quad \sum_i F_i(D_i(t)/\tilde{w}(t)) - \sum_j \lambda_j Y_j(t) \\ - \sum_j [Y_j(t) - K_j \tilde{a}_j(t)] \mu_j^Y(t) \\ + \theta(t) [\sum_j Y_j(t) - L(t) - \sum_i D_i[t, \tilde{w}(t), p_i(t)]] \\ \text{s. t. } Y_j(t) \geq 0 \quad Y_j$$

where:

- E_t = Expectation operator based on all information available to the utility when it makes decisions at t .
- $\mu_j^Y(t)$ = Dual variable on unit j 's capacity and availability
- $\theta(t)$ = Dual variable on meeting another unit of total demand

This Lagrangian has duality conditions

$$(3.1.7) \quad \mu_j^Y(t) > 0 \quad \text{--->} \quad Y_j(t) = K_j \tilde{a}_j(t) \quad (\text{Generator } j \text{ is fully loaded}) \\ \theta(t) > 0 \quad \text{In all cases of practical importance}$$

and equations 3.1.3 to 3.1.5 must hold exactly. Notice that $\theta(t)$ is the social value of another unit of energy generated at time t .^{1.6}

Information and Control: The achievable level of welfare will depend on how much information the central utility has and on how much it can control. Under full spot pricing, the utility can set a different $p_i(t)$ at each period $t = 1, 2, \dots$. Therefore (3.1.6) can be solved independently at each period. I will assume that the utility has full information and control of all generators: it knows $\tilde{a}_j(t)$ and sets $Y_j(t)$ for all j . But the utility can never completely know $\tilde{w}(t)$ or the value added functions F_i of all its customers. However under full spot prices it does not need this information; it needs to know only the current demand as a function of price: $D_i(t/p_i(t))$. As long as periods are short, and $\tilde{w}(t)$ is changing slowly it is sufficient to know last period's demand $D_i(t-1/p_i(t-1))$ and the effect of small changes from last period's price. This can be estimated from past behavior at the same time of day. Of course forecasts may be in error for individual customers, but the error in total demand will be quite small if periods are short. We will see later that customer specific demands have a comparatively small effect on prices; total demands are more important under most circumstances.

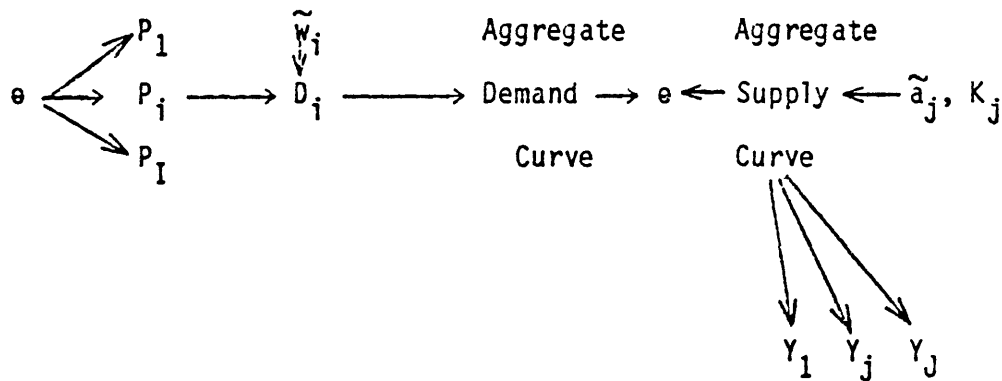
Rationing: Rationing of customers under full spot pricing is never necessary or desirable. The energy balance constraint (3.1.4) can be met by raising prices to reduce demands, if total demand would otherwise exceed total generating capacity.

Model Solution

The utility must simultaneously set prices $p_i(t)$ and generation levels $Y_j(t)$. That is, it must choose prices which balance supply

and demand, and set the level of supply. This appears complex because there is a different price for each customer. It will turn out, however that the dual variable $\theta(t)$ can be interpreted as a "spatially averaged price," and finding its value is the key to the solution.

I will first show how all the $p_i(t)$ are functions of $\theta(t)$. Thus total demand is a function of the single price, $\theta(t)$. I will then show the utility's supply curve as a function of $\theta(t)$. The intersection of these two curves at any time gives the equilibrium θ at that time. From there the individual $p_i(t)$ and $Y_j(t)$ are uniquely determined. Diagrammatically, this can be shown as follows:



To find optimal $p_i^*(t)$ given $\theta(t)$, differentiate the Lagrangian (3.1.6) with respect to i 's demand D_i to give the first order condition which an optimal price must satisfy:

$$(3.1.8) \quad \frac{\partial F_i}{\partial D_i} - \theta \left[1 + \frac{\partial L}{\partial D_i} \right] = 0$$

Substituting in customer behavior (3.1.2b) and losses (3.1.3) gives:

$$(3.1.9) \quad p_i^*(t) = \theta \left[1 + \frac{\partial L(t)}{\partial D_i} \right] = \theta \left[1 + 2 \underline{e}_i' \underline{B} \underline{D}(t) \right]$$

where \underline{e}_i = a vector of 0's with a 1 in the i'th position.

$\underline{D}(t)$ = vector of all demands.

To interpret (3.1.9), recall that the shadow price θ is the value of an extra kilowatt hour at the point of generation. But to deliver an additional kwh to customer i may require the utility to generate more than 1 kwh, since line losses may increase. Customer i 's effect on losses depends on its location and on $\underline{D}(t)$. Hence its price depends on these also. This will be discussed further below.

Under normal conditions $\partial L / \partial D_i$ is on the order of ten percent, and therefore optimal full spot prices are approximately equal to the "average" price $\theta(t)$. Prices also contain a term for incremental losses caused by the customer. Both terms will change stochastically over time.

Given a price, each customer chooses its demand $D_i[t, \tilde{w}, p_i(t)]$. These demands determine losses, and hence determine what total generation $G(t)$ must be to satisfy the energy balance (3.1.4). Given $G(t)$ the utility must set optimal dispatching level $Y_i(t)$. I will now show how optimal dispatching patterns are determined to meet different levels of $G(t)$.

Solving for optimal dispatching levels is easy since I have assumed each generator has the same effect on system losses. Given the unit availabilities $\tilde{a}_j(t)$ we can draw the short run system

marginal cost curve in Figure 3.1.1, which is in effect a short run supply curve. This enables us to visualize the process of finding the optimal $Y_j(t)$ in the Lagrangian (3.1.6).

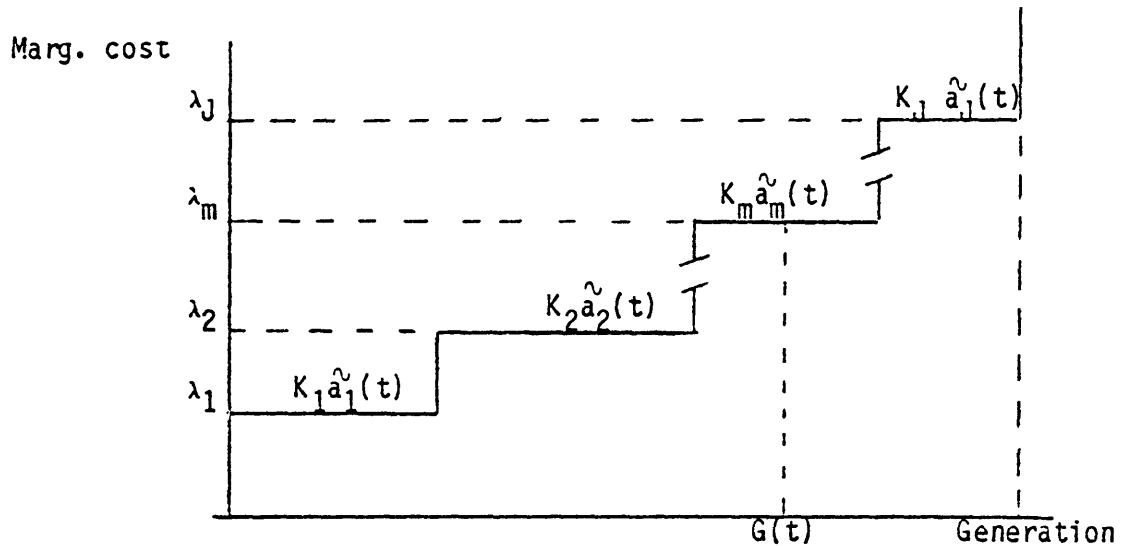


Figure 3.1.1

System dispatch curve at time t

The $Y_j(t)$ are optimally selected by dispatching units in ascending order of λ_j , until their total output equals $G(t)$:

$$(3.1.10) \quad \sum_{j=1}^{m(t)} a_j(t) K_j = G(t) \text{ and } m = m(t) \text{ is the "marginal unit" at } t.$$

The optimal dispatch pattern is then:

$$(3.1.11) \quad Y_j(t) = \tilde{a}_j(t) K_j \quad j < m(t) \quad \text{Fully dispatch units below } m$$

$$Y_j(t) = 0 \quad j > m(t) \quad \text{Do not dispatch units above } m$$

$$0 \leq Y_m(t) < \tilde{a}_m(t) K_m \quad \text{Partially dispatch } m$$

In Figure 3.1.1 all units to the left of $G(t)$ should be dispatched fully; marginal unit m should be partially dispatched as shown.

The dual variables on capacity are quite useful. Differentiate (3.1.6) with respect to each unit which is active at t . This gives the shadow value of a unit of capacity of type j :

$$(3.1.12) \quad \mu_m^Y(t) = \begin{cases} \theta(t) - \lambda_j & j = 1, \dots, m \\ 0 & j = m+1, \dots, J \end{cases}$$

Value of capacity of type j = (Value of a unit of energy) - (Cost of generating that energy using capacity of type j)

Define the system lambda $\lambda(t)$ as the marginal running cost of the last unit loaded at t :

$$(3.1.13) \quad \lambda(t) = \lambda_m(t)$$

System lambda is the short run marginal cost of energy at the generation point. Then there are two possible cases: either total losses plus demand for electricity is below the utility's available capacity, or it is not. The former case implies

$$Y_m(t) < \tilde{a}_m(t) K_m$$

which by (3.1.7) means $\mu_m^Y(t) = 0$. In this case of extra capacity available, the price parameter $\theta(t)$ is just the system lambda:

$$(3.1.14) \quad \theta(t) = \lambda_m = \lambda(t) \quad \text{if } Y_m(t) < \tilde{a}_m(t) K_m.$$

Note that $\theta(t)$ and $\lambda(t)$ are non-decreasing functions of total generation $G(t)$, regardless of whether there are long-run economies of scale for generation. They are non-increasing function of the (stochastic availability of generating units.

It is also possible that demand requires the use of all available generation capacity. Hence $m(t) = J$, $\mu_m^Y(t) > 0$, and by (3.1.12), $\theta(t)$ is above system lambda:

$$(3.1.15) \quad \theta(t) = \lambda_J + \mu_J^Y(t)$$

Define the curtailment premium:

$$(3.1.16) \quad \mu(t) = \theta(t) - \lambda(t)$$

$$= \begin{cases} 0 & \text{if not all capacity is in use.} \\ \theta(t) - \lambda_J & \text{if the last generating unit is fully loaded.} \end{cases}$$

or: The price parameter = System Lambda + Curtailment Premium

$$(3.1.17) \quad \theta(t) = \lambda(t) + \mu(t)$$

Discussion of Basic Results

We can now interpret the equilibrium value of the price parameter $\theta(t)$. Unit capacity and availability trace out an upward sloping instantaneous supply curve like Figure 3.1.1. Stochastic effects $\tilde{w}(t)$ and price responsiveness lead to a particular downward sloping instantaneous demand curve as a function of θ . It is the sum of the individual demands $D_i(\tilde{w}, p_i(\theta))$. The intersection of the instantaneous demand and supply curves gives the unique optimal $\theta(t)$, as shown in Figure 3.1.2.

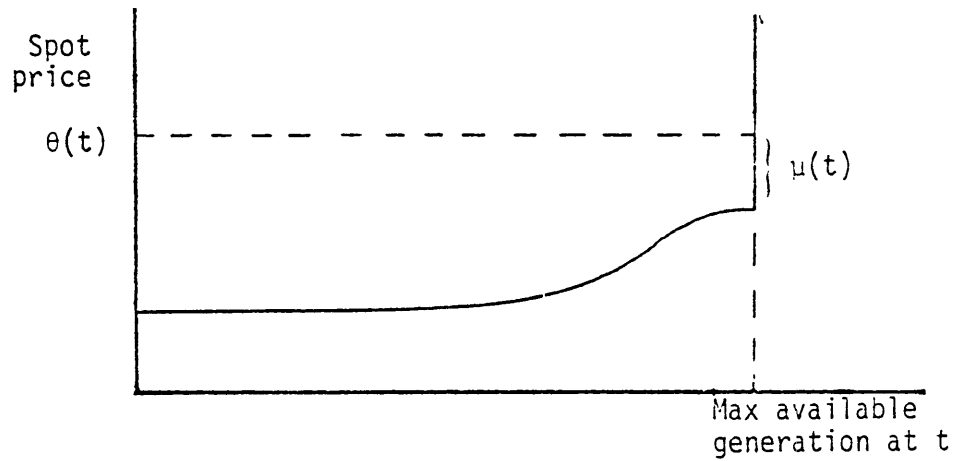


Figure 3.1.2

Equilibrating Instantaneous Supply and Demand

Under normal conditions $\theta(t)$ equals the system lambda, the short run marginal generating cost (fuel plus variable maintenance costs). When demand is very high or unit availability low, an additional curtailment premium $\mu(t)$ must be added as shown in Figure 3.1.2:

$$\theta(t) = \lambda(t) + \mu(t)$$

$\theta(t)$ corresponds to the value of electricity at the point of (marginal) generation. Customers which are located elsewhere on the T and D system see different prices, composed of the basic price plus a charge for their effect on system losses. From (3.1.9) we get:

$$(3.1.18) \quad p_i^*(t) = [\lambda(t) + \mu(t)] [1 + 2 \underline{e}_i' \underline{B} \underline{D}(t)]$$

A more accurate version of this key equation will be presented in Section 3.2. The \underline{B} matrix is a non-sparse matrix which depends on the strength and configuration of the T and D system. A variant of it is routinely calculated by utilities.

Implications

Several results about optimal prices emerge from equation 3.1.18 and its derivation.

- o These full spot prices are sufficient to achieve social welfare maximizing behavior by all customers, provided that they act as pure price takers. Predetermined prices cannot give this behavior at all times, and therefore cannot be preferable to full spot prices, except because of transactions costs. (See Sections 3.5 and 5.3.)
- o Customers with time-varying demands pay according to full spot prices at the time they use electricity, independent of earlier demands. Demand charges (typically based on maximum use during the previous month) are not needed to achieve the socially optimal demand.^{1.7}
- o Full spot prices are the same for a net user or net generator if they are at the same location. There is no difference between the equation for a cogenerator and the equation for a conventional customer.
- o A customer may increase or decrease incremental losses, hence have a price above or below $e(t)$.
- o The spatial component of full spot prices essentially depends on demands at every point on the system.^{1.8}
- o The difference in price between the two customers depends on demands elsewhere in the system.

- o The spatial term multiplier $e'_i B_{ij} D(t)$ is proportional to total demand, for equal increases in all demands. That is, doubling all demands will double the spatial price differences.
- o The full spot price to a customer depends on its own demand. First, increasing its demand moves the entire system up the instantaneous supply curve, increasing $\lambda(t)$ or $\mu(t)$. Except possibly when $\mu(t)$ is positive, i.e. total demand is constrained, this effect will be insignificant for all but the very largest customers or very small utility systems (by U.S. standards). Second, by altering its demand a customer alters line flows through the network and this changes its incremental effect on losses. Thus

$$\frac{\partial p_i^*(t)}{\partial D_i(t)} = 2[\theta(t)] B_{ij} \neq 0$$

where B_{ij} is positive and large if i is in a weak portion of the T and D system. If this effect is very large customers will not act as pure price takers, and will demand less than the socially optimal amount. Similarly, independent generators will withhold some output at times when they are on the margin. For small customers and generators these effects will be very small. See also the discussion of decentralized operation in Section 3.2.

Optimal Investment

The previous derivations were for a given utility capital stock. But in the long run the utility can build new generating units (and transmission lines). Each new unit shifts the instantaneous supply curve outward, potentially leading to lower short-run generating costs, lower spot prices, or lower curtailment. To the extent the new unit does this it will increase short term welfare. If the expected net present value of its impact on short term welfare (consumer's plus producers' surplus for a fixed capital stock) is greater than the cost of building the unit, then the unit should be built.

I will show that this welfare maximizing criterion for investment has a natural interpretation in terms of spot prices. The Appendix has a more rigorous derivation and a more complete model, including transmission investments.

Define $WST(t) = WST(\underline{w}(t), \tilde{a}_1(t), \dots, \tilde{a}_J(t), K_1, \dots, K_J)$ as the short term welfare achievable as the optimal solution of the pricing/dispatching problem (3.1.6). Then long term welfare is given by the discounted expected value of $WST(t)$, minus the cost of constructing generators K_1, \dots, K_J . Following standard practise I will assume for the present:

- o All units are constructed at once.
- o Unit size is continuously variable and there are no unit level scale effects in costs or reliability. To build a unit of type j and size K_j costs $c_j K_j$.
- o All units have the same lifetime T , and net present value factors are built into the costs c_j .^{1.9}

Then the long-term maximization problem is to choose K_1, \dots, K_j to solve:

$$(3.1.19) \quad \max E_0 \int_0^T \text{WST}(t) dt - \sum_i K_i c_i$$

where

E_0 = Expectation based on knowledge when investments are chosen.

The first order conditions for this are determined by differentiating (3.1.19) and (3.1.6) to give:

$$(3.1.20) \quad c_i = E_0 \int_0^T \frac{\partial \text{WST}}{\partial K_i} dt = \int_0^T E_0 \mu_j^Y(t) \bar{a}_j(t) dt$$

But $\mu_j^Y(t)$ is exactly the net revenue per megawatt unit produced by a unit of type j , i.e. the difference between the full spot price to a generator, and marginal operating cost when the unit is operating (Equation 3.1.12). So $E[\mu_j^Y(t) \bar{a}_j(t)]$ is the expected net revenue which would be earned by j if it is treated as its own profit center and paid the full spot price $\theta(t)$ for everything it generates.

Thus the optimal investment rule (3.1.20) is:

Build units of type j up to the point that the expected net revenue (under full spot pricing) of the last unit built equals the construction cost of the last unit.

This rule can be easily visualized by means of a price duration curve, which shows the cumulative probability of different spot prices, and is analagous to a load duration curve. For a fixed utility capital stock, the unfolding of time and random variables will give rise to

Figure 3.1.3
Spot Price over Time

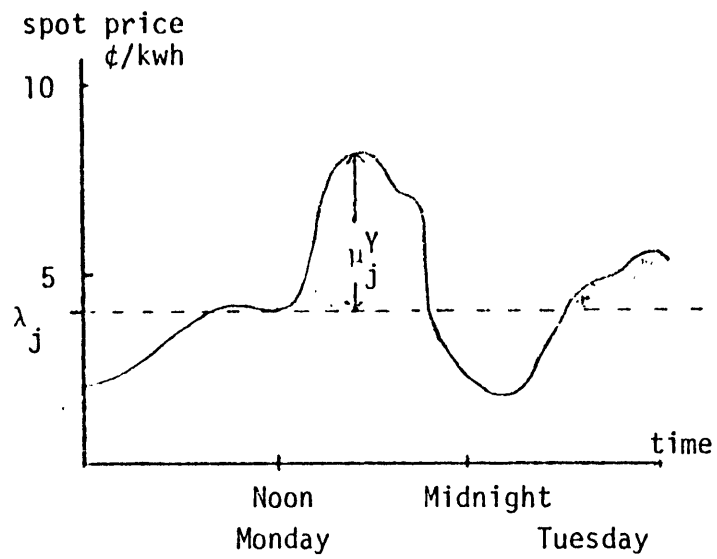
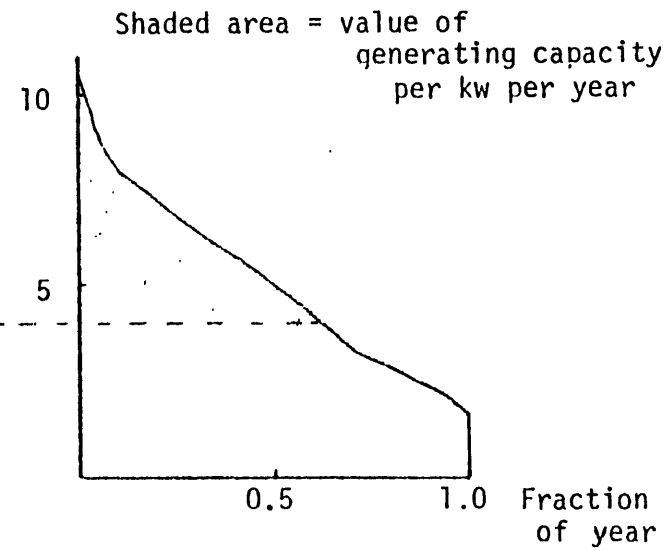


Figure 3.1.4
Price Duration Curve



full spot prices over time such as in Figure 3.1.3. Taking the probability distribution of $\theta(t)$ over the life of the plant gives Figure 3.1.4, which shows the fraction of time with each price or lower. (Chapter 5 presents an approximate price duration curve for a real utility.) For a unit of type j , $\mu_j^Y(t)$ is the vertical distance from λ_j to $\theta(t)$, as shown for Monday noon in Figure 3.1.3. The integral in (3.1.20) is then the shaded area above λ_j on the price duration curve, times the mean unit availability $E a_j$. (A normalization for the length of the interval T is also needed. If $\theta(t)$ and $a_j(t)$ are not independent then this is an oversimplification. The price duration curve should be drawn conditional on unit j being available. For example for solar units, it would be drawn only for prices at times when the sun is shining.) Hence if the area above the unit's marginal operating cost under the price duration curve exceeds the marginal capital cost of enlarging the units, then the unit should be enlarged.^{1.10}

Notice that the relevant area is much larger for a base load unit (one with relatively low λ_j) than a peaking unit. At the optimum, the shape of the price duration curve will be such that these different areas are exactly proportional to the relative capital costs (divided by expected availability) of each type of unit. Changing fuel prices will affect the shape of the curve, and the profitability of some units. For example an increase in coal prices would move the curve upward for those values of θ such that coal is the marginal fuel. This will affect the profitability of units which are below

coal in the loading order (nuclear, solar), but not of units which are higher in the loading order.^{1.11}

We now have a method of comparing the value of different kinds of units, with different reliabilities and different marginal operating costs. Units with lower availability can deliver less energy per megawatt of nameplate capacity. It is nonetheless conceivable that a unit with lower average availability

$$E \int_0^T \tilde{a}_j(t) dt$$

could be worth more than a unit with the same marginal operating costs and higher average availability. This will occur if the first unit is available more at times of high spot prices. For example compare a cogenerator which generates electricity whenever its plant needs steam, with a run-of-river hydro generator whose "availability" is determined by rainfall. The relative value per MW of capacity in each unit depends on the correlation of full spot prices with rainfall versus correlation of full spot prices with the cogeneration plant's steam demand. Models which arbitrarily penalize units for "low reliability" will miss effects like this.

All the above analysis of optimal investment criteria is correct only to the extent that the unit in question will be dispatched optimally, i.e., according to rule (3.1.11). If not, the investment must have social value less than $E \mu_j^Y(t) \tilde{a}_j(t)$. Fortunately, optimal operation can be achieved even without central utility ownership, as I will discuss in the next section. In Chapter 5 I will

show what happens to the value of investments when they are not dispatched optimally.

Conclusion

This section has presented a simple model of optimal spot pricing and investment by a welfare maximizing utility. Some of the specific results will be modified in the more realistic model discussed next. The basic results, however, are quite robust.

- o Optimal full spot prices vary stochastically as demand and supply (unit availability) fluctuate.
- o The price at each instant varies depending on where a customer is located. It does not depend on
 - whether the customer is a net buyer or a net seller
 - past or future demand by the customer (demand charges or price of reliability).
- o Full spot prices equal system lambda plus a demand curtailment premium, multiplied by an adjustment for line losses.
 - The curtailment premium $\mu(t)$ is normally zero, but can rise quite high if needed to prevent rationing.
 - The line loss factor is different at each point in the T and D network, and depends on the present spatial pattern of demand.
 - An additional component of prices will be introduced in the next section. It is usually small.

- o Socially optimal investment criteria are to build an addition to a unit if that addition, treated as its own profit center and paid the full spot price for whatever it generates, would be profitable.
 - Profitability depends on future spot prices, which are influenced by fuel prices and trends. This implies long run uncertainty, which affects the expected value of units.
 - The "price duration curve" is a convenient way to estimate expected profitability.
 - Changes in the shape of the curve can occur because of changes in fuel prices, the capital stock of generating units, or trends in demands. For example increased customer responsiveness to spot prices will flatten the curve, lowering the profitability of peaking units.

3.2 Extending the Basic Model

In this section I discuss extensions of the basic model of Section 3.1. Figure 3.2.1 shows the additional complexities covered in the Appendix which were not included in the basic model. It also shows how each issue is handled in conventional time-of-use pricing models, which will be reviewed in Section 3.4. Four of the extensions to the basic model have significant implications, and will be emphasized here. They are:

- o The joint use of spot and partially predetermined prices, for different customers.
- o The accompanying possibility that rationing will be needed.
- o Interperiod demand and supply effects. For example, customers can "store" electricity embodied in intermediate products.
- o Constraints on the flows and voltages in the T and D system. These influence spatial variation in spot prices.

In conclusion I will sketch some practical issues concerning the calculation and use of spot pricing and alternative methods of controlling an electric power system. Spot pricing, at least in theory, suggests the possibility of efficiently mixing regulated and unregulated competitive power generators. The basic model included the possibility that some "customers" were selling small amounts to the utility. Here I will discuss sales by much larger dedicated central station generating units.

Figure 3.2.1

Issues Covered in Different Models

Simple Model (Section 3.1)	Full Model (Appendix)	Traditional Model (Section 3.4)
oOnly spot prices	*Some customers on each pricing system	* Only predetermined prices
oRationing not necessary	*Rationing of customers on predetermined prices may be necessary	* Rationing necessary
oGeneral stochastic demand	Same	Simple stochastic demand
oStochastic unit availability	Stochastic unit and transmission availability and fuel/operating costs	* Deterministic unit availability
oConstant marginal operating costs	U shaped marginal operating costs	Constant marginal operating costs
oIntertemporal independence of supply and demand	* Cross-period demand and supply effects, such as storage	Intertemporal independence of supply and demand
oLine losses only	* Line losses plus constraints	*No line losses or T and D system
oMinor decentralized generation; all utility owned generation in one place	Partially or fully decentralized generation; ties with neighbors; spatially dispersed main generators	o No decentralized generation
oCustomers disaggregated	Customers disaggregated and have own capital stock	Customers aggregated
oAll investments at one time	Sequential investment	All investment at one time
oOnly real power; no reactive power	Same	Same
oZero transactions cost	Implicit transactions costs only	Zero transactions costs

*These differences from the simple model are important to the theoretical results.

Correct Full Spot Prices

In an actual utility the generating and transmission technology specifications in Section 3.1 must be extended in several ways. First, generators as well as customers will be spatially dispersed. Therefore incremental losses $\partial L(t)/\partial Y_j(t)$ may be different than $\partial L(t)/\partial Y_k(t)$. The effect of this is just like the effect of spatially dispersed customers: there is a different value (full spot price) for a kilowatt hour produced by generators at different points.

Second, the assumption of constant short run marginal costs, λ_j , is too simple. In fact generating units have U-shaped marginal heat rates, giving them upward sloping short run marginal costs as they near full rated output. Furthermore, large generating units are severely constrained in how fast they can change output without damaging the equipment. This means that marginal generating costs at t depend on output at $t-1$ and $t+1$.

Other important examples of intertemporal effects are centralized pumped storage of electricity, or decentralized customer storage of electrical energy in another form. Either type of storage will, if properly dispatched, act to partially level prices over the course of a day. For example the spot price during an afternoon demand peak will depend partly on how much storage was filled up earlier. Storage behavior is discussed extensively in Chapter 4.

The existence of the intertemporal effects makes the price duration curve a pedagogical tool rather than a rigorous analytical device. However it is still valid for some kinds of units (gas

turbines and others which I call "pure shutdown" in Chapter 4) and a useful way to think about most others.

Third, an actual T and D system has voltage magnitude and power flow constraints. That is, no more than its rated capacity can flow over a line or through a transformer. Since the flow over individual lines is determined by Kirchoff's Laws rather than direct central dispatching, this may constrain permissible generation/load configurations many miles away from the affected line. Define $\underline{Z}(t)$ to be the vector of the line flows and voltage deviations at every line and bus (node) in the T and D system. The Appendix shows how constraints on \underline{Z} lead to dual variables $\underline{\mu}^Z(t)$ and $\underline{\eta}^Z(t)$, which are non-zero if and only if the corresponding constraint is about to be violated.

With these additional complications, the optimal full spot price formula becomes:

$$(3.2.1) \quad p_j^*(t) = \left[1 + \frac{\partial L(t)}{\partial D_j(t)} \right] [\lambda(t) + \mu(t)] - \frac{\partial Z(t)}{\partial D_j(s)} [\underline{\eta}^Z(t) - \underline{\mu}^Z(t)]$$

The full spot price to j is the system λ adjusted for j 's impact on system losses and on system line constraints (which may be positive or negative, depending on j 's "location" in the TD network and the state of the TD system), plus the μ term to curtail demand if total system capacity is fully used. Here system lambda is slightly redefined, but basically it is still the marginal generating cost of the marginal unit in the generating order.

Equation 3.2.1 differs from the previous formula (3.1.18) by the

last term, involving T and D system limits. For a solidly built T and D system this term will rarely be large. Also, incremental transmission losses $\partial L / \partial D_j$ depend on the T and D network configuration, which is stochastic because transmission lines can be randomly knocked out.

Finally, $\lambda(t)$ can now depend on events before and after t . For example demand on many systems rises rapidly early in the morning. Large and efficient generating units may not be able to track this rise, necessitating the temporary use of less efficient units. This will raise the system lambda faster than the simple model would indicate.^{2.1}

Spatial Pricing: The Two Area Case

In conventional spatial models, the price difference between two points at one time must be less than or equal to the cost of shipping the good from one to the other.^{2.2} This arises from the standard transportation model, in which a feasible activity is to transport a unit of product from i to j , at a constant cost per ton mile. Electricity does not fit this model since there is no monetary short run transport cost, there are losses "in transit", and a fixed network is required. Finally the product cannot be allocated to a specific point; it is physically meaningless to talk about a specific generator j selling to a specific customer i , unless the two are connected only to each other and not to a T and D system.

For all these reasons, spatial price difference in an electric

power system under full spot pricing are complex. For one thing, the difference in price between two points depends on demands elsewhere in the system. For example, as line flows throughout the T and D network decline, so will spatial price differences. These points were discussed in general terms in Section 3.1.

Here I will consider an analytically tractable special case: two almost independent areas connected by a single transmission line. For example each area may be a separate utility, connected by a single tie line. If both utilities are using full spot pricing they will automatically interchange optimally along the tie line, and spot prices on both systems will be coordinated as if they were a single utility. We can solve analytically for the difference in spot prices at each end of the tie line.

Figure 3.2.2 taken from Elgerd [1971, Section 8-3] shows the two area problem at one instant. (The time argument will be suppressed for brevity.) D_i and Y_i are demand and generation at the i th bus for $i=1$ or 2 . Z_{12} is the power flow between them, and in this

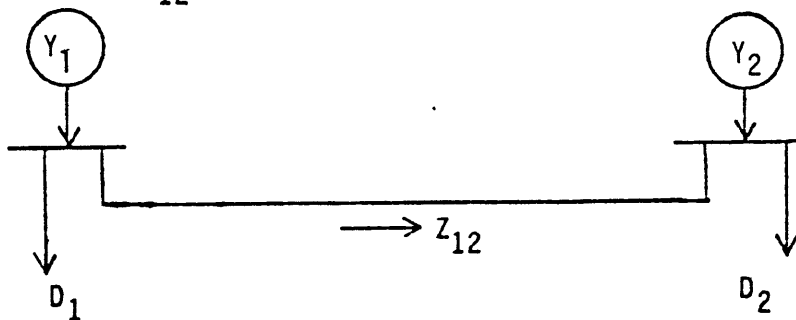


Figure 3.2.2

The Two Area Problem

example will be a flow from bus 1 to bus 2. It is approximately equal to $Y_1 - D_1$, the net generation at bus 1. L is the loss along the line, which must satisfy the energy balance equation:

$$(3.2.2) \quad L = Y_1 - D_1 + Y_2 - D_2 .$$

We know from (3.2.1) that the price difference $p_2^* - p_1^*$ is given by:

$$(3.2.3) \quad p_2^* - p_1^* = \left[\frac{\partial L}{\partial D_2} - \frac{\partial L}{\partial D_1} \right] [\lambda + \mu] + \left[\frac{\partial Z}{\partial D_1} - \frac{\partial Z}{\partial D_2} \right] [\underline{n}Z - \underline{\mu}Z]$$

= Differential effect on losses x value of losses
+

Differential effect on constraints

In the two-bus case, losses are proportional to the square of the tie line flow:

$$(3.2.4) \quad L = \Omega (Z_{12})^2$$

where Ω is a function of the resistance of the tie line.^{2.3} Then the differential effect on losses is:

$$(3.2.5) \quad \frac{\partial L}{\partial D_2} - \frac{\partial L}{\partial D_1} = 2 \Omega Z_{12} = 2 \frac{L}{Z_{12}} = \text{Twice } \underline{\text{average line loss}}^{\text{2.4}}$$

To evaluate the price difference (3.2.3) there are three cases, distinguished by the tautness of the T and D system constraints (the Z's). In general, the more heavily loaded the tie line, the greater the price difference between its two ends.

Case 1: No T and D constraints are active. So the second term of

(3.2.3) is zero and from (3.2.5) we get:

$$(3.2.6) \quad p_2^* - p_1^* = 2 \frac{L}{Z_{12}} [\lambda + \mu]$$

Thus the price at bus 1 is lower, and the percentage difference is twice the average loss on the tie line. For example if losses on the line at time t are 3 percent, then

$$p_1^*(t) = .97 [\lambda(t) + \mu(t)] = p_2^*(t) - .06 [\lambda(t) + \mu(t)]$$

Thus prices are about 6 percent higher at the receiving bus.

Case 2: The tie line between the two buses is fully loaded. In this case equation 3.2.6 sets a lower bound on the price difference. Prices at bus 1 will fall and at bus 2 will rise, as much as necessary to force D_2 to "back off" or Y_2 to increase, preventing demand/generation patterns which would give $Z_{12}(t) > Z_{12}^{\max}$. In effect each side of the tie line becomes an autonomous market, except that they have a flow of exactly Z_{12}^{\max} from bus 1 to bus 2.

In summary, under normal conditions the price difference between two areas connected by a single tie line will be approximately equal to the "non-spatial price" $[\lambda(t) + \mu(t)]$ times twice the average losses in the tie line. Average losses are proportional to the power flow in the line. So if no energy is flowing over the line, prices will be equal. Otherwise price will be lower in the "source" area than the "destination" area.

General Networks

I have discussed Figure 3.2.2 as if it is a complete system. But the "demands" D_1 and D_2 can arise from anywhere, including a complete T and D network with its own demands and generators. This causes two possible complications: the two buses may be connected by several pathways, and there may be active T and D constraints elsewhere in the system. Corresponding to cases 1 and 2 we therefore have:

Case 3: No T and D constraints are active, and the buses are connected by multiple pathways. (Which may be very indirect.) Then Case 1 provides a useful upper bound on the price difference. Choose any one pathway, no arc of which is fully loaded, and compute the (algebraic) average loss along it, as in the two bus case. Then the largest possible price difference between the two points is given by (3.2.6) where L/Z_{12} is the average loss along that pathway alone, as if the rest of the network did not exist. That is, the price difference is twice the average loss along this pathway, times $\theta(t)$.

Since other pathways do exist, the actual price difference will usually be lower. Multiple pathways can be mathematically combined to form an "equivalent circuit" which is then analyzed by the above procedure. In most networks at most times, a few pathways will dominate, and considering only them will give a good approximation to the true price difference. Notice that this difference still depends on power flows on all pathways which connect the two points, and therefore depends on demands throughout the utility system. Customers

whose neighbors have large positive demand (relative to the strength of the T and D network in their locale) will tend to see higher full spot prices, even if their own demands are small.

Case 4: T and D constraints are active, anywhere in the network. Then \underline{n}^Z or $\underline{\mu}^Z$ are nonzero and the second term of (3.2.3) contributes to the price difference. I have not been able to prove any general results about the term. Both absolute and relative prices will be affected. The price difference between the two areas can be increased, sometimes drastically. This can happen abruptly if a heavily loaded tie line is knocked out.

Rationing and Non-Spot Customers

Optimal full spot prices $p_j^*(t)$ can be calculated in real time for each customer, at least approximately. However, because of transactions costs it will not be desirable to have all customers on full spot pricing as will be discussed in Section 3.5. Non-spot customers may have to be rationed, if their unrationed demand/generation level would lead to violation of constraints on line flows, voltages, or energy balance.

The Appendix derives optimal rationing rules. The basic idea is that electricity used by the customer has a social value of $p_j^*(t)$ but a private value to the customer equal to its predetermined price $p_j^i(t)$. When $p_j^*(t)$ is much higher than $p_j^i(t)$ it may be

socially desirable to cut the customer off, even though this will disrupt its operations. In practice this will be appropriate mainly when the curtailment premium $\mu(t)$ reaches levels close to the average disruption cost of rationing a group of non-spot customers. Rationing these customers is then socially preferable to making customers on full spot pricing voluntarily curtail further in response to still higher spot prices.

Thus under optimal utility behavior:

- o The possibility of rationing effectively puts an upper bound on spot prices, equal to the marginal disruption losses caused by rationing.
- o The more participants are on spot pricing, the less often rationing will be needed for other participants, since the more likely that demand can be held down at spot prices below the disruption cost of rationing.
- o The probability that j will be rationed is an increasing function of $p_j^*(t) - p_j'(t)$. In particular, multiple rate classes may exist with different rules for updating their prices. (See Section 3.5.) All else equal, participants on infrequently updated prices will be rationed most often, since their prices will have the largest forecast errors.

One special case corresponds to present utility operation. If all participants are on the same non-spot prices then when rationing is needed the curtailment premium $\mu(t)$ jumps from zero to the social loss due to rotating blackouts (or whatever involuntary demand reduction

method the utility uses.) All of the previous formulas for full spot pricing and for investment still hold, even though spot prices are not actually being used.

Decentralized Operation and Investment

The theory of spot pricing was presented in Section 3.1 for a utility which owns and operates the T and D system and all but a few small generators. It also applies to situations where independent competitors own and operate a large amount of generation. Spot prices are calculated by the same formulas as before, and act as signals to generators to adjust their output levels in response to changing supply and demand conditions. If the generating firm is a perfect price taker, full spot prices lead it to self-dispatch exactly as if it were centrally owned, i.e. according to equation 3.1.11. The social value of a generation expansion for unit i (the right hand side of equation 3.1.20), is also the expected private profitability of the expansion if i is independently owned and paid optimal full spot prices at all times. Thus, to a first approximation, competitive generating firms under full spot pricing would behave as if owned by a welfare maximizing monopolist. Thus full spot pricing can, at least in theory, replace economies of scale due to unified ownership of generation.

The same logic applies to customer investments. Predetermined prices give lower incentives than full spot prices for many

investments. For example many customers can transform and "store" electrical energy as thermal energy or embodied in intermediate products. At present some utilities are trying to identify and subsidize such investments. Full spot pricing makes this unnecessary by internalizing to customers the social value of investments. Thus it is a method for utilities to encourage optimal investment and generation by customers, without exerting direct control or spending any money.

Naturally, to the extent that perfect competition by generators does not exist in an electricity spot market, behavior of independent firms will deviate from social welfare maximizing behavior even if properly calculated full spot prices are used. There are at least four possible deviations of a spot market from a frictionlessly competitive ideal.

- o The remaining central utility has strong market power, even if it is confined to calculating full spot prices and building and controlling the T and D system. While supply and demand forces will determine $\theta(t) = \lambda(t) + \mu(t)$ at each instant, a central utility could reconfigure or underbuild the T and D system to increase spatial price differences and its net revenues. Without thorough auditing it can also simply miscalculate prices, as long as it does so in a way which maintains the energy balance constraint.^{2.5} Of course this is not fundamentally different than the problem of controlling the behavior of a traditional utility using marginal cost rates, which will be discussed in

Section 3.6. Full spot pricing with decentralized ownership of generation does not eliminate the need for regulating the owner of the T and D system.

- o Individual generating firms might own enough capacity in a region to affect the system's λ at certain times. This type of market power is traditionally dealt with by antitrust action.
- o As discussed above each generator or customer will have some spatial market power. That is, a^2_{Lj}/aD_j^2 is nonzero, and therefore ap^*_{j}/aD_j will be nonzero (Equation 3.1.9). The magnitude of this effect depends on the strength of the T and D system.
- o Economies of scale in unit capital costs can lead to construction of units large enough to affect local prices; private investors will then size new units slightly below the social optimum. They will also retard construction of new units in the face of growing demand. This is discussed in the Appendix.^{2.6}

The above problems occur to some extent in many unregulated U.S. markets which have lumpy investment and non-zero transport costs. But the feasibility and desirability of fully decentralized ownership of electricity generation has other potential problems, such as the need for accurate real-time competitive market clearing. Some of these are discussed in the Appendix; others in Bohn et al. [1982]. The purpose of the discussion here is mainly to point out the possibility of a mixed system of central utility and competitive ownership of generators, and the need to use full spot prices to achieve efficient

coordination in such a system.

Conclusion

Optimal full spot prices vary over time, space, and state of nature as necessary to:

- o Maintain total demand at less than or equal to total current generating capacity, despite generator outages and demand fluctuations.
- o Maintain line limits and voltages within acceptable limits.
- o Equate the marginal cost to the marginal value of each kwh of energy for all users and all independent generators at all times.
- o Allow for the differential impact of each participant on line losses. These impacts change over time as total line flows change.

This leads to a full spot price equal to current short run marginal cost $\lambda(t)$ plus a "curtailment premium" $\mu(t)$ plus various individual-specific terms. The individual-specific terms are normally small compared with the other terms. All of these terms depend on the current generating capital stock, which may or may not be "optimal."

If all customers are charged full spot prices and are aware of the current price at all times, it will never be necessary or optimal to ration consumers, as any necessary level of demand curtailment can be obtained by raising spot prices. If some customers are not responsive to the current spot price it may be optimal to ration them. The

probability that a customer should be rationed is an increasing function of the difference between its full spot price and the actual price it is paying. Whenever that difference crosses the cost of disruption due to rationing, the customer should be rationed. The more participants are on spot pricing, the less often this will occur.

Optimal capital investment decisions have a natural interpretation in terms of full spot prices, even if such prices are not used. An investment in a plant should be made if that plant will have positive expected profits (net revenue minus capital costs) when paid full spot prices. (This condition is sufficient but not necessary; the necessary condition is weaker for plants large enough to influence the current spot price, i.e. with significant spatial market power.) Profitability of a new plant can be approximated by looking at a conditional price duration curve, which gives the cumulative probability of different prices. The net revenue of the plant is proportional to the area under the price duration curve above its short run marginal operating cost. Of course the number and nature of participants on full spot prices will alter the level of those prices and therefore alter the optimal generating mix. Switching participants from predetermined to spot prices will flatten the price duration curve, moving the optimal generating mix toward baseload units and away from peaking units.

The proper use of full spot prices to customers and decentralized generators provides closed loop feedback to help control the power system. It therefore makes the system more robust against

uncertainty, including demand fluctuations, outages, and long run forecast errors which lead to an ex post suboptimal capital stock.

But this is not achieved costlessly. Routine transactions costs for metering and for communicating prices are higher, the closer prices are to full spot. Therefore it will be optimal to have some participants on rates in which prices change more slowly. Such rates are discussed in Section 3.3. Optimal and practical assignment of participants to various rates is discussed in Section 3.5.

Furthermore, spot prices can increase the methods for opportunistic behavior by a profit maximizing regulated utility, possibly leading to a net social welfare loss. This problem is discussed in Section 3.6.

3.3 Optimal Predetermined Prices

The full spot prices derived and discussed above are "optimal" only with no transactions costs. When transactions costs are considered, it will generally be preferable to aggregate prices across time, across states of nature, and across space/participants. Different rates, with their own amount of aggregation, may be optimal for different participants. This section discusses optimal prices for a given level of aggregation along each dimension. For example, suppose we exogenously specify that prices can only change twice a day, and must be set each December for the following year. How should prices for the year be set? The rates discussed here are "optimal" rates for a given level of aggregation, under the assumptions that:

- o Only "first best" welfare issues are considered. Revenue constraints are ignored.
- o If multiple rates, and hence multiple prices at one instant, are used, no arbitrage selling is permitted between participants on different rates.
- o Assignment of participants to a rate is mandatory, not voluntary.

These assumptions are discussed and relaxed somewhat in Sections 3.5 and 3.6.

Given these assumptions, deriving exact equations for optimal aggregated prices is tedious but straightforward and not fundamentally new. Ellis [1981, Section IV.D] has a general exposition of the two-price case. Others have derived various special cases for

aggregation across one dimension.^{3.1}

I will present exact formulas for several special cases, then discuss the general cases. Assume for simplicity that demands by customer j are independent across time, i.e.

$$(3.3.1) \quad \partial D_j / \partial p_j(t) = 0 \text{ for all } s \neq t$$

Suppose that a single price must be chosen at time s , which will be in effect from t_1 to t_2 . Then the optimal level of this price is:

$$(3.3.2) \quad p_j(t_1) = p_j(t_1+1) = \dots = p_j(t_2) =$$

$$= \frac{E_s \left[\sum_{t=t_1}^{t=t_2} p_j^*(t) \times \frac{\partial D_j(t)}{\partial p_j(t)} \right]}{E_s \left[\sum_t \frac{\partial D_j(t)}{\partial p_j(t)} \right]}$$

where $p_j^*(t)$ = optimal full spot price (from Section 3.2)

E_s = Expectation based on information available at s .

Each term of the numerator of this expression is approximately equal to the mathematical expectation of the optimal full spot price, times the expected demand responsiveness:

$$(3.3.3) \quad E_s \left[p_j^*(t) \frac{\partial D_j(t)}{\partial p_j(t)} \right] = [E_s p_j^*(t)] \left[E \frac{\partial D_j(t)}{\partial p_j(t)} \right] - \text{COV}_s(t)$$

where $\text{COV}_s(t)$ = Covariance of $p_j^*(t)$ with $\partial D_j(t) / \partial p_j(t)$.

Thus $[E p_j^*(t)] - p_j(t)$ has the opposite sign as the covariance

term. Note that the covariance does not measure the coincidence of demand and spot price, but rather that of demand responsiveness and spot price.

If the rate requires presetting of prices but no time aggregation, and if the responsiveness of demand to price is uncorrelated with the full spot price, and if losses/T and D capacity limits are ignored, then (3.3.2) reduces to:

$$(3.3.4) \quad p_j(t) = E_S p^*_j(t) = E_S [\lambda(t) + \mu(t)]$$

This is the familiar result that "[optimal predetermined] price in each period must equal the conditional expected [short run] marginal operating plus rationing costs." [Crew and Kleindorfer, 1979 p 75] Analogous results hold for aggregation across time [Joskow, 1976 p 202] and space [Craven, 1974] in deterministic models. This result generalizes to: "Each optimal predetermined price equals the expected value of optimal full spot prices, averaged over the time periods and geographic areas where the price will be in effect."

Discussion

These price equations, and more general versions which relax equation 3.3.1, have the following properties.

- o Optimal prices are weighted averages of the optimal full spot prices.
- o The weights are larger for demand/generation which is more responsive to price, and would be zero for any totally unresponsive demand or generation.

- o The weights themselves may be stochastic, and the full spot prices are always stochastic. Therefore optimal aggregated prices are the expected value of a weighted average.
- o If behavior is not independent across time, then optimal aggregated prices at one time will depend partly on the full spot prices at other times, weighted by the cross-responsiveness of demand.
- o The responsiveness of demand to price may depend on when the prices are revealed. It is a non-decreasing function of the amount of advance warning received. (See Chapter 4.) Therefore the notation $\partial D_j(t)/\partial p_j(t)$ is misleading, as this quantity can depend on the entire structure of rates and anticipated future prices.
- o There are obvious informational problems in evaluating even equation 3.3.2, much less more accurate equations which incorporate $\partial D_j(t)/\partial p_j(s)$. In practice, it may be appropriate to ignore the covariance terms, and set predetermined prices equal to unweighted average spot prices. This will be done in the case studies of Chapter 5.
- o Switching some participants to another rate may change their behavior, therefore the level of full spot prices, therefore the level of aggregated prices. But it does not change the formulas by which these are calculated, nor the optimal investment equations. Instead those equations are evaluated at different points. Therefore the insights of Sections 3.1 and 3.2 still apply, even to systems with no participants on full spot

pricing. This will be useful in the next section.

- o Aggregation across space/customers is analogous to aggregation across time. When full spot prices are different for different customers on the same rate, more weight should be placed on getting the correct price for customers who are most price-responsive. (Equation 3.3.2 with summation across customers as well as across time.)

3.4 Comparison with other Public Utility Pricing Models

How does spot pricing differ from conventional prescriptions for public utility pricing and investment? The idea of time differentiated prices goes back at least to 1949.^{4.1} Until Brown and Johnson [1969] the models were purely static and deterministic. During the 1970's various authors presented prescriptions for time-of-use pricing in static models with demand uncertainty. Their analysis can be considerably simplified and generalized by using the concept of state contingent spot prices, i.e., spot pricing. I will first discuss conventional time-of-use pricing models. I then discuss models with dynamic investment, differential reliability and spatial pricing. I conclude with previous authors' work on spot pricing.

Time of Use Pricing

The "standard" time-of-use pricing models are surveyed in Gellerson and Grosskopf [1980] and Crew and Kleindorfer [1979].^{4.2} They include Wenders [1976], Crew and Kleindorfer [1976, 1979 Ch. 4 and 5], Turvey and Anderson [1977, Ch. 14], and various predecessors. These models include multiple types of generators and stochastic demand, but in other ways are even simpler than the model of Section 3.1, as was shown in Figure 3.2.1. I will comment on their most important limitations, then discuss models which address some of those limitations.

- o Generating unit availability is modeled by simply derating unit sizes at all times. This fails to properly penalize large units, and it gives inaccurate estimates of the probability that rationing will be needed. It also gives no guidance for how to evaluate new technologies such as solar and cogeneration, whose "availabilities" are correlated with demands by other customers.^{4.3}
- o There is no analysis of how or when prices should be recalculated. These models rule out frequent recalculations (by spot pricing) by assumption. By assuming infinitely repetitive demand cycles and stable factor prices they show no need for annual or less frequent recalculations. Demand and cost trends are thus not considered.
- o Like Section 3.1, these models treat all investment as occurring at once. Investment is really a sequential process. True utilities never have the static optimal capital stock of these models, because conditions change more rapidly than capital stock turns over. Therefore pricing equations which assume optimal capital stock, i.e. assume that short run and long run marginal costs are equal, have limited practical value. In fact long run marginal costs can only be calculated conditional on a particular scenario or probability distribution of demand and factor prices. This problem is addressed by Ellis [1981], discussed below.

- o Like Section 3.1, the models assume that demands and generating costs are independent from one hour to another. This is very convenient, since it allows the use of single load duration curve (or price duration curve). Nonetheless the availability of storage [Nguyen, 1976] or demand rescheduling can have a major impact on optimal prices and investment policies.
- o The models ignore transmission, which is equivalent to assuming an infinitely strong transmission system. This is not feasible when setting practical rates for power buybacks, but these models give no insight into how to price over space. Current debates about "wheeling tariffs" indicate the importance of this issue when trying to encourage independent generation by firms located in the territory of a monopolistic utility.
- o The models do not use the device of state contingent prices. Therefore, the investment conditions derived in the models are hard to interpret, although they are correct (given the limiting assumptions above). For example, Crew and Kleindorfer [1979, p. 77] interpret their results only for the case of interchanging units which are adjacent in the loading order. Littlechild [1972] showed the way out of this problem, but his point was apparently missed by subsequent authors.

All of the above limitations are dealt with in the full model of this chapter, which was discussed in Section 3.2. Other earlier models also deal with them individually, and in several cases illuminate particular issues better than I have done.

Dynamic Pricing/Investment Models

Several authors present deterministic explicitly dynamic models which can be interpreted as deterministic versions of spot pricing. Crew and Kleindorfer [1979, Ch. 7] give a continuous time optimal control model with one type of capital. They get the result that:

Whatever the level of capacity, price is to be set to maximize instantaneous [short run] welfare returns subject to the given capacity restriction. [That is,] price should equal SRMC. Of course, at optimum capital stock is adjusted so as to equate SRMC and LRMC....In the event of ... a fall in demand, [optimal] price is less than LRMC, then capacity would be allowed to decline until equality between price and LRMC were re-established.[p 113]

They are thinking here on a time scale of years, not hours; they reject continuous adjustment of prices to reflect the actual level of demand. Nonetheless, their model can be interpreted in terms of hourly price adjustments.^{4.4}

Turvey and Anderson [1978, Ch. 17] have a discrete time dynamic model which leads to discontinuous prices, as capital investment is made in lumps. However they reject this approach: "It is apparent that, for one reason or another, such fluctuations are unacceptable." They also acknowledge that investment decisions must be made before price decisions, and with more uncertainty about future demands, but they do not incorporate this into their models. [p 305]

Ellis [1981] explicitly models sequential investment and pricing decisions. He concludes that welfare optimal pricing rules differ according to whether prices must be set either before or after investment decisions are made. He uses dynamic programming to look at how the character of optimal sequential investments depends on capital stock irreversibility and the sequential revelation of information about future demands.

Spatial Pricing

Several previous authors have studied how public utility prices should vary over space. Relevant models include Takayama and Judge [1971] (which was not directed at electricity), Craven [1974], Dansby [1980], Scherer [1976, 1977], and Schuler and Hobbs [1981]. All of these models are deterministic and most are static. Only Scherer has an accurate model of electricity line losses and line constraints, or includes T and D investment options.

Scherer's mixed integer programming model of an electricity generation and transmission network is an excellent deterministic version of Section 3.1's model. In his model spatially distinct prices appear as dual variables on demand at each point in the network. In his numerical case study he found that prices between different points at the same time varied by up to 30 percent. The absolute and percentage variations across space changed over time. [1977, p 265ff] He does not discuss these results, but presumably they reflect the different losses resulting from different optimal

load flows at each level of total system demand, as discussed in Section 3.1.

Much of Takayama and Judge concerns pricing across space. They consider only competitive markets, but use an explicit optimization method of finding equilibrium, so their analysis is equally applicable to a welfare maximizing monopolist. They assume a constant transport cost per unit between two points, no transport capacity limit, and no losses. This makes their models more appropriate for conventional commodities than for public utility products such as electricity. They also assume linear demand and supply functions. But their framework does provide insights into more general spatial and temporal pricing problems. For example they discuss "no arbitrage" conditions which bound the price differences between different locations.[1971, p 405] Their models do not include capital, so they provide no insights into optimal investments in transport facilities.

Pricing of Reliability

One way to view spot pricing is that it allows customers to choose their own reliability levels. Marchand [1974] has a model in which customers select and pay for different reliability. The utility allocates shortages accordingly, when curtailment is necessary. His approach differs from (and is, except for transactions costs, inferior to) spot pricing because customers must contract in advance, and therefore have no real time control over their level of service. Also, customers not curtailed by the utility have no incentive to adjust demands.^{4.5}

A simple version of Marchand's proposal is in use in the U.S. and elsewhere. Called "direct load control", it involves the utility turning off specific equipment of the customer's. Despite its increasing use [Morgan and Talukdar, 1979; Gorzelnik, 1982] optimal pricing and use of direct load control has not been studied by economists.^{4.6, 4.7}

Spot Pricing

State contingent pricing of public utility services was apparently first proposed by Vickrey, under the name "responsive pricing". His original article [1971] presented a general discussion using as examples mainly long distance telephones and airlines. The emphasis is on curtailment premia, rather than on marginal production cost changes over time. Later manuscripts on electricity develop the ideas in more detail, including some discussion of optimal investment criteria [Vickrey, 1978 p 12], metering requirements and designs, pricing of reactive energy, and short run marginal operating costs (system λ). He proposes that utilities be free to set prices however they want over time, subject only to limits on total profits similar to those discussed in Section 3.6 under "Equal Revenue Rates."^{4.8}

Vickrey's essential insight was that prices can be set after some random variables are observed, and optimal prices should reflect this. Since his original article different versions of this basic idea have been developed independently and under different names, with

varying levels of rigor. These include:

- o "State preference" approach to pricing electricity. [Littlechild, 1972] A formal stochastic model of both pricing and investment under static conditions. Both operating costs and capacity constraints are modeled, but with homogeneous fixed coefficient technology, i.e. only one kind of capital.
- o "Time varying congestion tolls" for a highway or communications network. [Agnew, 1973; 1977] A formal deterministic optimal control model incorporating only capacity constraints and delays. No discussion of investment.
- o "Spot pricing" of electricity. [Schweppe, 1978; Schweppe et al 1980] Subsequently developed into the formal model of Chapter 3's appendix. [Bohn et al 1981; Caramanis et al 1982]
- o "Real time pricing" of electricity. [Rand, 1979] Informal; no specific proposal.
- o "Load adaptive pricing" of electricity. [Luh et al, 1982] A game theoretic model; nonlinear prices allowed. Quadratic production costs assumed, with no capacity constraints and no investment. Their formulation allows for games between one utility and one consumer which is not a pure price taker.
- o "Flexible pricing" of electricity. [Kepner and Reinbergs, 1980] Informal.

Many other authors have explicitly rejected the idea that prices can be set after events are revealed. For example, Crew and

Kleindorfer [1980, p 55] write: "For the case of the regulator setting the price ex post, he or she would either have to allow a market-clearing price or have some deliberate arrangement for setting the price above or below the market clearing price. Were the regulator [to allow] the market clearing price, he would, in effect, be giving up his right to regulate price." Turvey and Anderson [1977, p 298] are even more adamant in their rejection of spot pricing:

...for a wide class of random disturbances (but not for all), it is not possible to respond to the resultant random excess or shortage of capacity by adjusting prices. Failure of a generating plant on Thursday cannot be followed by a higher price on Friday, and the price in January cannot be raised when it becomes apparent that January is colder than usual. Even though telecontrol makes the necessary metering technically possible, it would be expensive, and... there would be difficulties in informing consumers of the new price. It would also be scarcely possible to estimate its market clearing level. Sudden and random price fluctuations would in any case impose considerable costs and irritations on consumers. Hence responsive pricing that always restrains demand to capacity is not practicable, and some interruptions are thus desirable. Their rejection thus appears to be based on the belief that the transactions costs of spot pricing would outweigh any possible benefits.

3.5 Assigning Participants to Different Rates

This section discusses a common and important problem for any public utility. Which pricing systems (rates) should be offered? How should different customers be assigned to them? The preceding portion of this chapter derived "optimal" rates assuming zero transactions costs. The resulting prices were called full spot prices. But as Chapter 2 showed, different systems for changing prices will have different transactions costs. Therefore the optimal pricing system for a participant depends on the characteristics of that participant, the stochastic and deterministic rates of change of the optimal spot prices, and the transactions costs of different pricing methods. Rates which are closer to full spot prices should be offered to the most price-responsive participants and on systems where those prices fluctuate the most.

This section makes several points about how customers should be assigned to different rates.

- o The social welfare maximizing rate for each customer depends on the customer's size and how it would behave under various rates, and on the transactions costs of different rates.
- o Any rate other than full spot pricing can create a subsidy, that is, a wedge between private and social costs. This subsidy can be positive or negative and is customer specific. It must be made up by the utility or other customers. Therefore which rate a customer is on affects profit distribution as well as total social welfare.

- o Therefore, customers will not always voluntarily choose the socially preferred rate for themselves.
- o The utility cannot adjust rates so that "on average" customers will self assign to the socially preferred rate or one close to it. The problem is analagous to what happens in competitive insurance markets with adverse selection: those receiving large positive subsidies under a rate drive everyone else off that rate.
- o Mandatory assignment of customers to rates, which is standard practise for some public utilities, cannot be done optimally either. Such assignment would require unobservable customer specific information.
- o In practise a combination of mandatory and voluntary assignment will probably give "reasonably good" results, and is the best that can be done.

Quantitative illustrations of these points will be given in Chapter 5 for selected "case study" customers. The rest of this section develops the points more rigorously and in more detail.^{5.1}

Social and Private Optimal Assignment Criteria

Which rate a customer or independent generator is assigned to will affect three costs:

- o Communications and other transactions costs.
- o The value of electricity used by the customer in response to prices under the rate.

o The customer's value added as a result of its electricity use. (This was called F_j in Section 3.1.)^{5.2}

The social and private assignment criteria are both "assign the customer to the rate which maximizes its expected value added, minus transactions costs and the expected value of electricity used." This sum is the net social or private welfare gain under a rate. The difference between social and private criteria is that a profit maximizing customer will value electricity at its price under the rate in question, whereas the social value of the electricity used is always the full spot price at the moment of use. Under any rate except full spot pricing there will sometimes be a divergence between social and private value; therefore the customer will compare rates differently than will a social welfare maximizer.

I will give a small model which permits precise discussion. Let:

$p^*(t)$ = Full spot price at time t .

$p'(t)$ = Price at t under an alternate rate.

$D_j(t, p^*)$ = Customer j 's demand at time t if it is on full spot prices.

$D_j(t, p')$ = Customer j 's demand under the prices of the alternate rate.

$\underline{D}_j(p)$ = Vector notation for the above demands; one element each period.

$F_j(\underline{D}_j(p))$ = Customer j 's value added if it demands $D_j(t, p)$ at time t .^{5.3}

Customer j 's expected contribution to gross social welfare if it is under full spot prices p^* is therefore (see eq. 3.1.6):

$$(3.5.1) \quad W_j(p^*) = E_0 F_j[D_j(p^*)] - E_0 \int_0^T p^*(t) D_j(t, p^*) dt$$

where E_0 = Expectation based on information available when the assignment to a rate is made.

T = Problem horizon = time when customer can be reassigned to another rate.^{5.4}

But if customer j is under the alternate rate, its expected contribution is:^{5.5}

$$(3.5.2) \quad W_j(p') = E_0 F_j[D_j(p')] - E_0 \int_0^T p^*(t) D_j(t, p') dt$$

Thus the socially optimal assignment criterion is to compare the change in gross social welfare with the change in transactions costs, and to select full spot prices p^* if:

$$(3.5.3) \quad W_j(p^*) - W_j(p') > \text{Additional transactions costs under full spot prices, over and above transactions costs under rate } p'.$$

How does this compare with j 's own profit maximizing criterion for selecting between p^* and p' ? Its expected net revenue under p^* is exactly $W_j(p^*)$. But expected net revenue under p' is $W_j(p') + S_j(p')$, where the subsidy $S_j(p')$ is defined as:

$$(3.5.4) \quad S_j(p') = E_0 \int_0^T [p^*(t) - p'(t)] D_j(t, p') \quad dt$$

Of course full spot prices have no subsidy: $S_j(p^*) \equiv 0$. Thus if given the choice, customer j will select rate p^* if:

$$(3.5.5) \quad W_j(p^*) - W_j(p') > S_j(p') + \text{Additional transactions costs}$$

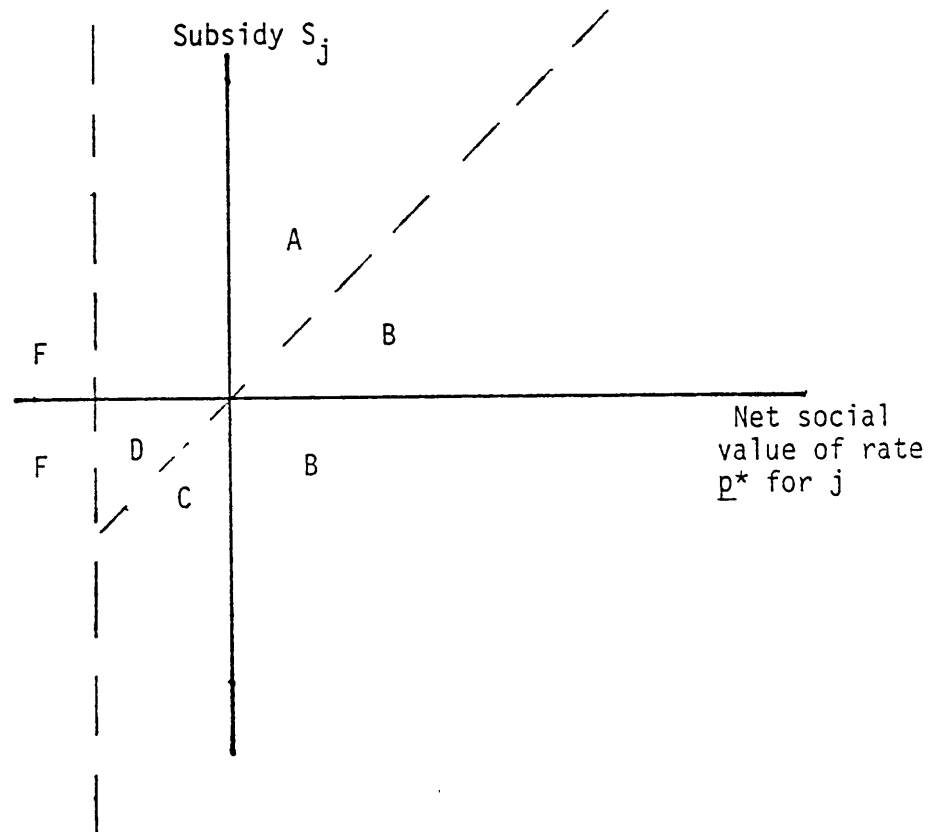
Thus the subsidy distorts private choice between the rates.^{5.6}

Figure 3.5.1 shows the private and social selection criteria graphically. Given a customer's value added function F_j and the two sets of prices p^* and p' we can determine the subsidy and gross social welfare under each rate. Any customer can then be represented by a point in Figure 5.3.1, with the vertical axis showing subsidy $S_j(p')$, and the horizontal axis showing the net improvement in social welfare due to spot prices, $W_j(p^*) - W_j(p') - \text{additional transactions costs}$.

The socially optimal assignment is to put j on full spot prices if it lies to the right of the vertical axis in Figure 3.5.1. But given a choice, j will select full spot prices only if it lies to the right of the 45° line in the Figure. Thus if j falls in region A or region C, it will not select the socially optimal rate.

Figure 3.5.1

Self Assignment Versus Socially Optimal Assignment



Region AUB = Region where socially optimal to be on full spot prices.

Region BUC = Region where j will choose full spot prices voluntarily.

Region A = Region where j will choose \underline{p}' though it is socially undesirable.

Region C = Region where j will choose \underline{p}^* though it is socially undesirable.

Region F = Infeasible.

Implications

From Figure 3.5.1 and the preceding equations we can derive several implications for comparing two rates, one of which may be full spot prices:

- o If customer j 's behavior will be the same on one rate as on the other, then the rate with the lower transactions costs is socially preferable for that customer.
- o The gross social welfare change of (3.5.3) will depend on the customer's size and responsiveness to spot prices. It will therefore be socially optimal to use more sophisticated pricing methods for customers which are larger or more responsive (in percentage of demand) to prices. This is consistent with the model of Section 2.5.
- o If two rates have the same transactions costs, the one which is closer to full spot prices should be used.
- o Whether a customer self-selects the socially optimal rate depends on its subsidy $S_j(p')$, which is a weighted average of the difference between p^* and p' , using $D(p')$ as the weights. The larger the absolute value of the subsidy, the less likely the customer is to select the socially desired rate. The subsidy has three components, and may be positive or negative.
 1. Correlation of stochastic demands with full spot prices.
 - 5.7 Customers with weather sensitive loads which are correlated with spot price will tend to have larger subsidies under any predetermined price than do other customers.

2. Correlation of the cyclic pattern of demand with time aggregation error in rate p' . Customers with weekday only demands will be subsidized by flat (non time-of-day) rates.
3. Deliberate differences between average prices under the two rates. For example $p'(t)$ may be deliberately set higher than $E p^*(t)$ in order to reduce the subsidies of all customers on rate p' .

Improving Self Assignment: The Adverse Selection Problem

The above suggests that by correcting for the subsidy $S_j(p')$ the utility can persuade customers to self select optimally. Unfortunately, the proper correction is participant specific, and attempts to do this have paradoxical properties.

If the utility actually calculates $S_j(p')$ for customer j and charges that as a lump sum for being on rate p' , this turns out to be exactly equivalent to putting j on full spot prices anyway. The cost to the customer of another unit of demand at time t is, from (3.5.4), $p'(t) + [\partial S_j(p') / \partial D_j(t)] = p^*(t)$.

A second approach is to charge some predetermined lump sum if the customer selects rate p' , and set this lump sum using a combination of individual data and aggregate data for all customers on the rate.^{5.8} This approach will often be easy to implement, but cannot give results which are accurate. Because of adverse selection it may lead to all but a few participants selecting the rate \underline{p}^* .

Suppose that the central utility can cheaply observe $\sum_j D_j(t)$

for all participants in a class which are on rate \underline{p}' . Then the utility can calculate $\sum_j S_j(\underline{p}')$ and charge each participant a share of the total class subsidy. For example, if each participant has its consumption measured once a month, then monthly consumption is a natural variable to use for shares:

$$(3.5.6) \text{ Charge to } j = \frac{\int_0^T D_j(t, p') dt}{\sum_j \int_0^T D_j(t, p') dt} \times \sum_j S_j(p')$$

where $T = 1$ month.

Unfortunately, a rule such as this will set off a process of adjustment such that too many participants will select the more sophisticated rate. Mitchell [1980] describes the process for the choice between two kinds of local telephone rates. Suppose that initially the total class subsidy is zero in equation (3.5.6), so the charge is zero. Then some participants with negative subsidies will fall in regions B or C in Figure 3.5.1, and will elect rate p^* . Total class subsidy of those remaining on the rate become positive, so the charge will be raised according to equation 3.5.6. This shifts the 45° line of Figure 3.5.1 to the left, shrinking region A and increasing the participants who choose p^* . Eventual equilibrium is established with only a few participants left on rate p' . These will be the participants which initially were heavily subsidized. Region C, participants choosing full spot prices even though the social gains do not outweigh the social costs, will be quite large. This is an example of adverse selection skewing the participants in a market.

Mandatory Assignment

Can the utility do a better job by mandatory assignment?

Historically this seems to have been assumed, at least for electric utility services. But in order to decide what rate customer j should be on, the utility should evaluate (3.5.3), which requires knowing something about how j would behave under alternate rates, and what the value of that change in behavior is to the customer. These will depend on the customer's options to substitute electricity for electricity at a different time and for other inputs to production. No central utility can know each customer's opportunity set. Even for classes of customers with many members, experimental methods will mainly give an indication of the mean and variance of changes in gross social welfare under different rates, which is not sufficient.^{5.9}

A reasonable approach to the assignment problem is therefore to use a mixture of mandatory and voluntary assignment. Participants can be divided into classes based on more-or-less exogenous characteristics, as is done today. Ownership of particular types of capital, such as electrical or thermal storage equipment, would be an important criterion for membership in some groups. Within each class participants might be offered a choice from among two or more rates, with the range of choices overlapping among different classes.

A typical set of rates offered under this approach might include:^{5.10}

- o "Full" spot pricing, with prices changing every 15 minutes to one hour. Mandatory for customers which already have hour by hour recording demand meters. Mandatory for some

participants with their own generation capacity. Voluntary for other large and medium sized customers.^{5.11}

- o Weekly time-of-day pricing. Each day is divided into several periods, with a repetitive cycle of prices each weekday. Once a week the prices for the next week might be recalculated, but the specification of intervals would remain the same.
- o Monthly time-of-day pricing. The same as weekly time-of-day pricing, except that meters would be read and prices recalculated only once a month.
- o Monthly flat pricing. Based on a conventional one-dial kilowatt hour meter. The price would be recalculated as often as the meters are read. In effect this system is in use today for most residential customers. It would be available on a voluntary basis for small electricity users (households); and not available at all for larger users.

It is important to remember that the optimal range of rates and "optimal" assignment rule will be utility-specific. This will be shown in Chapter 5.

How many rates to offer depends on the relative transactions costs and social welfare benefits of additional rates. Each new rate carries with it some transactions costs which are independent of the number of participants on that rate. If all these costs were zero, it would be optimal to have an infinite spectrum of rates. Instead, the additional transactions costs must be weighed against the improvement

in net social surplus for participants assigned to this rate instead of the previously available rates (equation 3.5.3). Craven [1974] discusses a crude procedure for determining the number of spatially differentiated rates. An additional rate will be more desirable the better the method for assigning participants to it.

3.6 Marginal Cost Pricing and Opportunism

Section 2.3 discussed the problem of opportunistic behavior by public utilities. There is a conflict between efficient price signals to customers, and avoiding opportunistic behavior by the utility. Marginal cost prices, whether they are full spot or predetermined, do not solve this conflict. Even if a hypothetical welfare maximizing utility in a particular market would exactly break even under marginal cost prices, a profit maximizing utility with the same rules for setting prices could increase its profits. For example it could underbuild and undermaintain some types of generating units, or buy more expensive fuel to increase its "marginal cost". This would shift the system dispatch curve to the left, and if done "properly" raise marginal cost more than average cost, increasing net revenues.

In practice, overt actions to raise costs deliberately, such as derating a unit at times of system peak, would not be likely. For one thing, they would require an explicit conspiracy within the firm which would be hard to keep secret. For another, some of them would go against the professional code of the utility's operators and engineers. Nonetheless, to allow incentives for opportunism is bad practice, and would raise the suspicions of customers. It would be very tempting for financially strained utilities to under-forecast load growth and therefore underbuild their system, given incentives like these. What can be done to remove such incentives while preserving marginal cost pricing, and particularly spot pricing?

The answer is, of course, to adjust prices in such a way that the

firm's total revenue approximately equals its total costs, almost regardless of its behavior. This in turn leads to two problems:

- o How to give the utility incentives to operate efficiently.
- o How to maintain efficient price signals to customers, i.e. price close to marginal cost even though marginal cost will rarely be close to average cost.

I have nothing to add to the discussion of the first problem, which was initiated by the Averch-Johnson literature and is surveyed in Schmalensee [1979]. I will give several partial solutions to the second problem. I will show that marginal cost prices, including full spot prices, are compatible with traditional rate of return regulation. Most of the benefits of full spot prices can be achieved within a traditional regulatory framework, even though that framework is based on accounting measures of average costs, rather than economic measures of marginal costs.

"Equal Revenue" Rates

The possibility of reconciling spot pricing with traditional average cost based regulation arises because the two are concerned with vastly different time scales. A utility's profits are determined by total costs and total revenues over an interval of a year or more, while much of the benefit of spot pricing comes from getting the right change in prices over the space of a few hours.

Suppose that a regulatory rule (such as a rate hearing plus fuel adjustments) determines the utility's allowed total revenues for the

next twelve months. Traditionally the next step is to estimate total demand over the period, and set a single flat price:^{6.1}

$$(3.6.1) \quad p^{\text{trad}}(t) = \frac{\text{Total Allowed Revenue}}{\text{Estimated Total Demand}} \quad \text{for } 0 \leq t \leq 1 \text{ year}$$

Full spot prices for this utility over the next year will oscillate above and below p^{trad} , probably crossing only about twice a day. To some extent these variations in full spot prices will tend to cancel each other. That is, the utility's actual total revenue under full spot prices may be quite close to the total allowed revenue, even though full spot prices are only rarely close to the traditional price at the same moment. The discrepancy between total annual revenues under the two rates must be less than the mean absolute deviation between the prices of the two rates.^{6.2}

Thus full spot prices can be adjusted to give the same total revenue over a year as traditional prices, and yet not be radically altered by the adjustment. For example suppose the adjustment is made by adding a constant amount Δ to full spot prices at each instant. Then Δ can be set as:

$$(3.6.2) \quad \Delta = E_0 \frac{\int_0^T [p^{\text{trad}}(t) - p^*(t)] D(t) dt}{\int_0^T D(t) dt}$$

where

$D(t)$ = Total demand of spot pricing customers at time t

T = Interval of averaging (e.g. one year)

E_0 = Expectation operator

and prices charged are then:

$$(3.6.3) \quad p^{\text{adj}}(t) = p^*(t) + \Delta$$

Equation 3.6.2 shows that the size of the adjustment Δ is determined by the discrepancy between the traditional price and the average full spot price over the period from 0 to T. This is a utility specific amount, and will depend on factors such as the regulatory treatment of construction in progress. For one class of customers and utility examined, Δ was about positive ten percent of traditional prices; that is, the utility's revenue would have been ten percent lower under full spot pricing. This compares with a standard deviation of full spot prices of about 51 percent of the mean. (Section 5.1) Thus for this utility the deviation from socially "optimal" prices due to the need to control opportunism is much less than the deviation due to arbitrarily keeping prices flat.

This is a simple explanation of a procedure which can be conducted with much more sophistication.

- o "A "rolling average" of the discrepancy between allowed and actual revenues can be used, instead of a fixed-horizon forward looking procedure.
- o The longer the interval of averaging, T, the better. For example when a new unit is completed, $p^{\text{trad}}(t)$ rises as it is put in the rate base, yet full spot prices will fall since the system's short run cost curve is shifted to the right.^{6.3} If T = 1 month, then Δ will have to increase drastically the month a new unit comes on line,

whereas if $T = 2$ years the discontinuity will be much less.

- o In general the size of Δ will depend on the difference between the utility's actual capital stock, and its optimal capital stock if it could start over and build according to the optimal investment equations of Section 3.1. It will also depend on regulatory treatment of the costs of debt and equity, regulatory definition of the rate base, the interval T , and other factors.
- o Proper discounting should be incorporated into equation 3.6.2.
- o Full spot prices can be adjusted in other ways than adding a constant Δ at all times. Also, the necessary total adjustment in revenues could be achieved by declining block rates, discrimination among customers, or other methods considered in the literature on "optimal deviations from marginal cost prices".^{6.4}
- o This general adjustment procedure is of course beneficial with time varying prices other than full spot prices, such as traditional time-of-use prices.
- o The utility's net revenues (gross revenues minus expenses) won't necessarily be the same under equal revenue rates as they would have been under traditional rates, since spot pricing will alter the demand profile. But if they wish, regulators can construct Δ to equalize net revenue instead of gross revenue.

Conclusion

Any system of marginal cost based prices, whether spot or predetermined, must be examined to see if it will aggravate opportunistic behavior by the utility. Calculating prices according to the formulas in Sections 3.1 through 3.3 provides some protection against price manipulation/price discrimination, and assurance of close to first best optimal incentives for customers/generators to whom they are applied. However a regulated utility can still manipulate prices by changing its operating and investment behavior.

Several cases exist for which such opportunism is not a problem.

- o A self-regulated, publicly owned utility can already manipulate prices; marginal cost or full spot pricing does not aggravate this ability.
- o Spot pricing by a utility which is buying more energy at the current spot price than it is selling at that price is not a problem. There are two sources from which a utility may be purchasing: independent generators or cogenerators within its territory, and neighboring utilities. But a given utility might switch from net spot purchaser to net spot seller each day. Only if the utility sells to few customers on spot prices, or owns little generating capacity, would it never be a net spot seller. This case does fit spot pricing buybacks under PURPA.^{6.5}

- o The utility could be required to treat revenue from spot customers as "negative purchases", and pass the revenue through to other customers through the purchased power clause. Obviously this cannot be done for all customers, and might distort the prices of non-spot customers.

For regulated utilities that apply marginal cost prices as a net seller, opportunism is a potential problem. The best that can be done is to give the utility the same incentives it has under conventional rates. This can be done via what I call "equal revenue rates". Such rates lead to prices which at each instant are close to optimal full spot prices, yet over time give the utility the same revenue as conventional rates.

The approach of "equal revenue rates" is very different than, but consistent with, the conventional analysis of marginal cost based prices for a profit constrained utility. The traditional analysis examines how to allocate a total revenue deficit or surplus among different customers, based on their elasticities or other criteria. The analysis is essentially static. My approach looks at how to smooth over time the revenue deficit or surplus each moment. Even after smoothing there may be some non-zero deficit or surplus, which can be allocated to different customers according to the usual procedures. A key insight is that the longer the averaging interval (T in equation 3.6.2), the smaller the expected value of the deficit or surplus will be, as a percentage of total revenue.

Equal revenue rates cannot eliminate Averch-Johnson or reverse A-J incentives for a utility whose traditional prices are based on a rate of return calculation. Spot pricing provides a potentially powerful tool for a utility to avoid new plant constructions, since spot prices help constrain demand to available supply. The same is true for predetermined time-of-use rates, but spot pricing is even more effective. Thus if a utility is consistently being allowed an inadequate rate of return on new units, it is conceivable that spot pricing could hurt its customers by helping the utility delay construction. Conversely, spot pricing also reduces the disruption caused by a utility which cannot or will not build new plants, since it leads to more efficient use of the electricity which is generated, and voluntary curtailments instead of involuntary rationing.

FOOTNOTES TO CHAPTER 3

0.1 All three have transmission constrained by existing network capacity. Electricity and natural gas lose a fraction of the amount transmitted for each mile traversed. Costs and losses for transmitting water are slightly different. Therefore the spatial pricing equations for water will be slightly different than those developed here.

1.1 This is accurate for household as long as they have approximately constant marginal utility of income. Similarly the assumption of risk neutrality is adequate if the cost of electricity is a small fraction of a firm's profits. This assumption is violated for generating firms. However, most of the risk of price fluctuations is diversifiable risk. Elsewhere I argue that fluctuations in annual profit will often be less under spot pricing than under predetermined prices.

1.2 The notation used here implicitly assumes that demands are independent each period; price at one time does not affect demand at other times. Since I have also assumed that short run generating costs are independent each period, it is mathematically correct to treat all change over time in the exogenous random variable $\tilde{w}(t)$, and solve the model for a general period as a function of \tilde{w} . This is a very standard approach, and leads to the use of price duration curves, discussed below. I will drop this assumption in Section 3.2 and the Appendix, in order to discuss storage.

1.3 Losses are directly determined by line flows, but line flows are in turn determined by demand and generation at each point. Thus losses also depend on the generation pattern, but by assuming all generators are at a single point in the T and D network their effect can be added into the loss matrix for customers. This simplifying assumption will be dropped in the appendix, leading to a different optimal price for each generator. Also, losses depend in part on reactive power flows, which are not discussed in this thesis.

1.4 Throughout this thesis I will assume that this constraint must be met exactly at all times. Other researchers are investigating the consequences of relaxing this assumption.

1.5 This objective function ignores income redistribution. See Schmalensee [1979, Chapter 2] for a defense of this approach on practical grounds. Also, the use of an expected value of consumers' surplus is criticized by Rogerson [1980] on the grounds that the marginal utility of income to households varies in response to the same exogenous random shocks which change consumers' surplus. Here, however, the "consumers" are other firms, not households, and consumers' surplus is the profits of these firms. By the standard

assumptions of the Capital Asset Pricing Model, expected profit maximization is the proper objective except for systematic risks. Furthermore, if long-term contracts are available, consuming and producing firms can lay off all risks; see Section 2.4.

1.6 This model is similar to previous models of electric power systems, except for stochastic unit availability, line losses, and its use of full spot pricing. See Section 3.4 for a more complete comparison with past work.

1.7 The model formulation did not allow for demand charge, so it might appear unfair to say they did not appear in the solution. However full spot prices alone were sufficient to lead to socially optimal demands. Demand charges penalize a customer's maximum demand regardless of when it occurs, which is suboptimal. It is conceivable that in a model with transactions costs, demand charges might be better than full spot prices for some small customers, since demand charges have lower meter costs. However only in very special cases would they dominate time-of-use rates.

1.8 The reason is that the B matrix is not sparse; demand anywhere influences load flows throughout the system, except in special cases.

1.9 The Appendix covers the realistic investment problem, with discounting, overlapping continuous stock and some analysis of unit level scale issues.

1.10 Of course this is only a partial solution method. Adding more capacity of type j lowers prices, hence lowers the price duration curve for all prices of λ_j or above. So the optimal amount of capital of each type depends on what other capital is being built.

1.11 When intertemporal effects are considered, a change in coal price can alter the shape of the entire price duration curve, hence affect the profitability of peaking units.

2.1 Utilities presently calculate the quantities $\lambda(s)$ and the B matrix which gives $\partial L(s)/\partial Y_j(s)$, for each of their major generators and in real time (see Elgerd [1971, Chapter 8]). The same algorithms and data base can be used to make this calculation at major load buses. Since these are usually the major components of full spot prices, this means that utilities can easily calculate spot prices under normal conditions.

2.2 Equality holds during periods when the good is being shipped.

2.3 Equation 3.2.4 is approximate since it does not include reactive power flows. It is more accurate when losses are small. A similar approximation works in more complex networks also: losses for the whole network are the sum of losses along each line, which are still approximately proportional to the square of flow on that line.

2.4 The derivation of equation 3.2.5 requires recognizing that, because of the energy balance constraint, increases in D_1 or D_2 must be accompanied by compensating changes elsewhere. For concreteness suppose the change is an increase in Y_1 . Then D_1 has no effect on losses since the flow from Y_1 to D_1 does not flow over the line. Also $\partial Z_{12}/\partial D_2$ is approximately 1. Using this and differentiating equation 3.2.4 with respect to Z_{12} gives equation 3.2.5. It also holds if we designate Y_2 as the responsive generator, or any linear combination of Y_1 and Y_2 .

It is very easy to double count in deriving (3.2.5). In fact Elgerd [1971] makes this mistake; he corrects it in the second edition, but uses a derivation much harder for a non-engineer to understand.

2.5 For example the utility could claim that a tie line to a major independent generator was on the verge of overloading, justifying a lower price to the generator and a higher price to customers. However the utility must manipulate prices so that demand and supply are both decreased by corresponding amounts or else the network will fail conspicuously. To detect manipulation, auditors need not worry about the average level θ of prices, but only about spatial price differences.

2.6 Eaton and Lipsey [1978] discuss the interaction of scale effects and spatial market power. The problem they present is in addition to those discussed here.

3.1 Craven [1974] treats aggregation across space. Joskow [1976, p 202] discusses aggregation across time. Caramanis et al [1982] cover aggregation across states of nature, i.e. predetermined prices.

4.1 Boiteux [1949], Steiner [1957]. See also Vickrey [1955].

4.2 Important literature not mentioned in either includes Marino [1978], Scherer [1976, 1977] and Vardi and Avi-Atzhak [1981].

4.3 Stochastic availability was introduced into the economics literature by Vardi et al [1977]. Their model, and its extension by Vardi and Avi-Itzhak [1981] is similar to those discussed here, except that it takes the level of system reliability as an exogenous design criterion.

4.4 Koenker [1977] has a continuous time deterministic model which he does interpret as leading to continuously varying prices over the course of a day. He follows the standard static assumptions of repetitive demand cycles and all capital chosen at once, so he does not deal with pricing when capital stock is non-optimal. Since his model is deterministic, spot pricing is not an issue.

4.5 A simpler version of Marchand's basic proposal is modeled by Tschirhart and Jen [1979] for a profit maximizing monopolist. The monopolist can control each customer's circuit breaker individually, and can price discriminate freely between customers. Under these assumptions the monopolist would have higher profits by using two-state spot pricing instead of a two-state circuit breaker, if such pricing could be implemented with no increase in metering costs.

4.6 Berg [1981] compares direct load control with time of use pricing. But he ignores the fact that direct load control is state contingent while TOU pricing is not. This severely limits the value of his analysis.

4.7 Dansby [1979] models a very simple version of direct load control, for a profit maximizing utility.

4.8 This approach has significant problems. It gives the utility complete freedom to set prices over space and time, subject only to a single constraint. Spot pricing as I model it places very tight constraints on prices; the utility must go to considerable trouble to manipulate prices, and has limited ability to change prices at one time and location without also changing them at others. See Section 3.6.

5.1 Apparently there has been little systematic study of how customers should be assigned to rates. Acton and Mitchell [1980] present an analysis of the choice between flat rates and simple time-of-use rates, for residential customers. My discussion follows their basic approach of comparing transactions costs with conventional welfare gains. Mitchell [1980] is also relevant and will be discussed below.

5.2 The discussion in this section applies equally to both customers and independent generators. As shown in the Appendix, the notation is completely general, keeping in mind that generators have negative "demands". For simplicity I will talk in terms of customers only.

5.3 This value added function is a scalar valued function of demands over a time interval. Only in the special case of "intertemporal independence", which was assumed in Section 3.1, does it decompose into the sum of value added at each moment. See the Appendix for a rigorous definition of value added. See Chapter 4 for models which show how to estimate the change in value added as a function of prices.

5.4 To avoid making this a dynamic programming problem I assume T must be set in advance.

5.5 Equation (3.5.2) uses the assumption that $D_j(t, p')$ - $D_j(t, p^*)$ is too small to change optimal spot prices significantly. This simplification is reasonable except for very large participants, as long as we are considering shifts of one participant at a time. When considering whether or not to offer a whole new rate with a large number of participants, however, the change in optimal spot prices as a result should also be considered.

5.6 For simplicity I will assume that all transactions costs are born by the end user. If not, the utility can internalize costs by charging a lump sum equal to its own incremental transactions costs. I am also ignoring rationing costs, which will affect social welfare and private profit equally.

5.7 These can be proven by expanding (3.5.4) around $E[p^*(t) - p'(t)]$.

5.8 A variant of this is to alter the prices $p'(t)$ for all customers in the class, more at some times than at others, to try to reduce the subsidy.

5.9 None of the existing time-of-use experiments will be sufficient to assess behavior under spot or partially spot prices, since the experiments have used fully predetermined prices.

5.10 Caramanis et al [1982] also proposed a spectrum of rates based on intuitive reasoning. Some of the rates proposed in that paper appear to be dominated.

5.11 This rate might have extensive spatial pricing, with participants on different substations and volages paying slightly different prices. Such spatial differentiation increases communications costs, however, and might be optimal only for extremely large participants.

6.1 This is an oversimplification. For example allowed revenues and prices may be calculated separately for each rate class. The procedure I will describe can be used to parallel the effects of any regulatory procedure.

6.2 This is true only if $p^*(t) = p^{\text{trad}}(t)$ at least once during the year. However the wide variation in $p^*(t)$ in virtually all U.S. utilities today guarantees this. The only possible exception today is in the Pacific Northwest. Even there, full spot prices at times of sustained heavy rainfall will be close to zero since water not used for generation must be wasted.

6.3 This assumes no Construction Work in Process.

6.4 Hall [1982] has proposed an interesting variant of lump sum adjustments in situations such as this of potential opportunism. His approach is applicable when customers are of very different sizes, and thus gets around the difficulty of assessing or paying each customer an equal "lump sum" fee when customers are of very different sizes. However, the transactions costs of his approach are not clear and might be quite high.

6.5 Utility interchange agreements today can also give incentives for opportunistic behavior whenever the utility is a net seller to neighbors.

APPENDIX TO CHAPTER 3
FORMAL DERIVATION OF FULL SPOT PRICES

Introduction

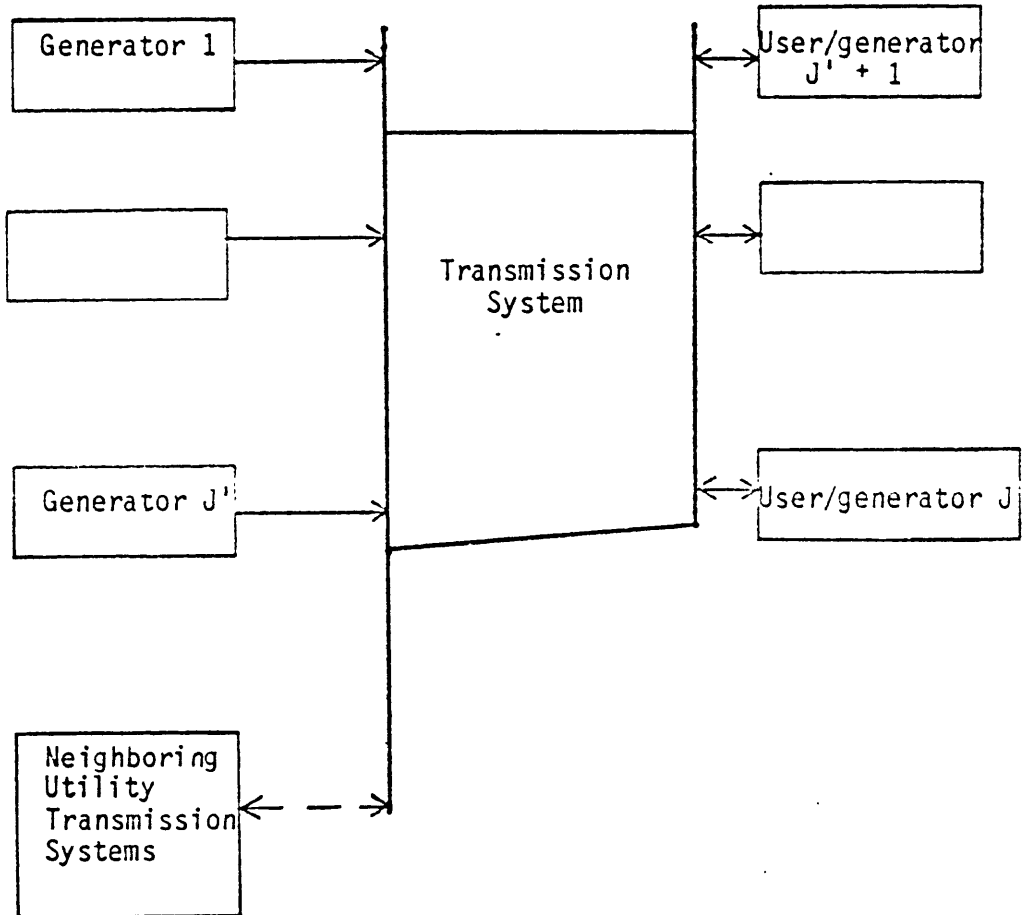
This appendix specifies the social welfare maximization problem for a monopolistically owned and operated firm. The model is more accurate than that of Section 3.1; Figure 3.2.1 showed the differences. Except for ignoring reactive power, this model includes all aspects of electricity production and use, including situations where not all customers are on spot prices. (Transactions costs were discussed in Section 3.5, and lead to not all customers on spot prices. I do not explicitly model transactions costs here, however.) The implications of the model were discussed in Section 3.2.

The underlying network is shown in Figure 3.A.1

Electricity participants, indexed by $j = 1, \dots, J$, are net users or net generators. All are interconnected by the transmission and distribution (TD) system which is owned and operated by the central utility. The utility also owns and operates some of the net generators.^{A.1} The first problem is to determine the socially optimal behavior of all participants. This can be divided into optimal short-run behavior given the capital stocks, and optimal long-run investment behavior. The second problem is to induce the independent (not centrally owned) participants to follow this behavior using only the limited control variables and information available to the central utility. The controls are of two kinds: prices and

Figure 3.A.1

Simplified Electric Power Network



rationing. Participants choose their electricity generation/demand levels in response to those prices subject to constraints imposed by rationing.

I now present the formal assumptions and notation. Individual participants are described first, then the transmission and distribution system. Finally social welfare maximization is defined.

Participants

In reality, a participant may be either a firm, a household, or a neighboring utility. We will model all participants as price-taking expected profit-maximizing firms.

Participant j is characterized by:

$Y_j(t)$ = electricity generated at t . Negative for net users.

\underline{K}_j = installed capital stock. For example, generator type and size.

F_j = $F_j(Y_j(t)/\underline{K}_j, \tilde{a}(t); t = 1, \dots, T)$ = value added by j .

Behavior of participants occurs over time. Only in simple cases will value added at each instant be independent of behavior in the past and future (see Chapter 4). Therefore value added F_j is defined over an interval, such as a day, week, or year. These intervals will be called cycles, and have periods $1, \dots, T$. Value added is the value of j 's production minus the cost of all variable inputs except electricity. Thus it can be thought of as a normalized production function for a technology with one input, electricity.^{A.2}

F_j depends on capital stock and on:

$\tilde{\underline{a}}(t)$ = Exogenous random variables. These include weather, outages, and prices of other products, especially electricity complements and substitutes. In particular define $\tilde{a}_j(t)$, $j = 1, \dots, J$ as 1 if firm j can operate at t , 0 if it is down for maintenance and cannot produce. Intermediate values are also possible, corresponding to partial derating. (In Section 3.1, I defined a separate weather variable, $\tilde{\underline{w}}(t)$, which influenced demand. Here there is no inherent distinction between demand and generation, and I have merged $\tilde{\underline{w}}$ into $\tilde{\underline{a}}$.

Some firms may be subjected to exogenously imposed rationing, which is "announced" after the participant has selected its desired generation for the period, $Y_j(t)$. I will assume such rationing is all or nothing for an individual participant.^{A.3}

Define $r_j(t)$ = 0 if j is not rationed at t
 = 1 if it is rationed.

Hence actual electricity generation/use at t is

$$Y_j^R(t) = [1 - r_j(t)]Y_j(t)$$

If rationing is imposed, j loses revenue or does not have to pay for electricity. It also suffers additional disruption effects (which may be quite severe) of:

$$R_j(r_j(t), Y_j(t)/K_j, \tilde{\underline{a}}(t); t = 1, \dots, T).$$

Rationing is at least as costly as voluntary curtailment. That is,

$$(3.A.1) \quad F_j(Y_j(t), Y_j(s)/\dots) \\
- R_j(r_j(s) = 1, r_j(t) = 0, t \neq s, Y_j(s), Y_j(t)/\dots) \\
\leq F_j(Y_j(t), Y_j(s) = 0/\dots).$$

This is guaranteed since the function F has within it the option of self-curtailment to 0 at time s . Also, since if no rationing is imposed during a cycle, no disruption is experienced,

$$R_j(r_j(1)=0, \dots, r_j(T)=0, Y_j(1), \dots, Y_j(T)/\dots) = 0.$$

This is convenient since it will turn out that optimal $r_j = 0$ for j on full spot pricing. Hence rationing costs can be ignored for these participants.

With these definitions, the net revenue of firm j over the cycle $1, \dots, T$, is given by:

$$(3.A.2) \quad NR_j(T/K_j) = F_j(Y_j(t)/K_j, \tilde{a}(t); t = 1, \dots, T) + \\
+ \sum_t p_j(t)Y_j(t) \\
- R_j(r_j(t), Y_j(t)/K_j, \tilde{a}(t); t = 1, \dots, T)$$

where $p_j(t)$ is the price of electricity to firm j at time j . Note that $p_j(t)Y_j(t)$ is negative for net users of electricity, and positive for net generators; the opposite is true for F_j . Thus the same specification holds for both. Net revenue corresponds to the familiar short run producer's or consumer's surplus.

For later use define

$$\emptyset = \text{set of all participants, } j = 1, \dots, J \\
\emptyset^+ = \text{set of all net generators}$$

ϕ^- = set of all net users.

I will assume in this chapter that behavior between cycles is independent, except for the fixed inputs \underline{K}_j which are durable. This allows a finite horizon model of electricity production and use decisions, $Y_j(t)$, $t = 1, \dots, T$.

Each firm's objective is to maximize its expected profits, subject to constraints on behavior. Expected total profits are given by:

$$(3.A.3) \quad \sum_{n=1}^{\infty} E \left[NR_j(nT/\underline{K}_j(nT)) - I_j(\dot{\underline{K}}_j(nT), \underline{K}_j(nT), nT) e^{-rnT} \right]$$

subject to

$$(3.A.4) \quad \underline{K}_j[(n+1)T] = \underline{K}_j(nT) + \dot{\underline{K}}_j(nT) \quad \text{Investment and}$$

$$(3.A.5) \quad Y_{\min,j}(t) \leq Y_j(t) \leq Y_{\max,j}(\underline{K}_j, \tilde{a}_j(t)) \quad \text{Production limits}$$

where:

r = discount rate^{A.4}

$\dot{\underline{K}}_j(nT)$ = net capital stock added during cycle n

E = expectation operator

$I_j(\dot{\underline{K}}_j(nT), \underline{K}_j(nT), nT)$ = investment cost during cycle of net

additions $\dot{\underline{K}}_j$ given initial capital stock \underline{K}_j .

$Y_{\max,j}(t)$ = maximum generation level. This will be a function of capital stock \underline{K}_j and of outages \tilde{a}_j . In particular if $a_j(t) = 0$, $Y_{\max,j}(t) \leq 0$; the unit cannot generate. (In Section 3.1, I used $Y_{\max,j} = a_j K_j$.)

$Y_{\min,j}(t)$ = minimum generation level at time t . Negative for net

users. Y_{\min} and Y_{\max} are important mainly for generators.

$Y_j(t)$ during the current cycle is assumed to have no effect on later cycles, or on investment costs.

This assumption of risk neutrality is a reasonable approximation for this problem, since spot prices will have roughly the same effect on the uncertainty of year-to-year profit fluctuation as do conventional rates, which are adjusted every year. (For example, average cost-based rates may be raised if total demand falls below projections. Mean spot prices would fall in this event. The exact comparison depends on how "conventional" rates are set, and on the use of adjustments to spot prices to reduce utility opportunism. The latter are discussed in Section 3.6.) See also footnote A.6.

The Transmission and Distribution System

All participants in an electric power system are connected by the transmission and distribution (T and D) system, as in Figure 3.2. Define \underline{k}^b as the capital stock of the T and D system: lines, transformers, and associated protection and control equipment. Each participant j is at a different point in the network.

Voltage magnitude and power flow constraints have to be imposed to prevent damage to the T and D network itself and to insure satisfactory operation of generation and usage devices. Define

$\underline{z}(t) = \underline{z}(Y_1^R(t), \dots, Y_j^R(t), \tilde{b}(t), \underline{k}^b)$: Vector of voltage deviations

from nominal design level at all network buses, and power flows through all lines and transformers at t . One element of \underline{Z} for each bus, each line, and each transformer.

These voltages and line flows depend on the injections $Y(t)$, random network events, $\tilde{b}(t)$, and of course the nature of the network itself, \underline{K}^b . The constraints are:

$$(3.A.7a) \quad \underline{Z}_{\min}(\tilde{b}(t), \underline{K}^b) \leq \underline{Z}(t) \leq \underline{Z}_{\max}(\tilde{b}(t), \underline{K}^b)$$

Note that $\underline{Z}_{\min} \leq 0$ and $\underline{Z}_{\max} \geq 0$. (These constraints sometimes involve several periods. This more general case is avoided to simplify notation.)

The T and D system has losses which depend on conditions throughout the T and D system. Define total real energy losses at t as:

$$L(t) = L(Y_1^R(t), \dots, Y_j^R(t), \tilde{b}(t), \underline{K}^b). \quad A.5$$

Conservation of energy dictates an energy balance constraint.

Define

$$(3.A.8a) \quad e(t) = \sum_j Y_j^R(t) - L(t)$$

$$(3.A.8b) \quad -e(t) = 0$$

In general, the stronger the T and D system (the larger \underline{K}^b), the lower $L(t)$ and $\underline{Z}(t)$. However, increasing \underline{K}^b has a cost. Define

$I^b(\underline{K}^b(nT), \underline{K}^b(nT), nT)$ = investment cost to increase T and D capital stock from \underline{K}^b to $\underline{K}^b + \dot{\underline{K}}^b$.

These costs are paid by the owner of the T and D system. Costs of operating the T and D system are essentially independent of the Y_j 's, and therefore their net present value is included in investment costs, I^b . Of course,

$$(3.A.9) \quad \underline{K}^b[(n+1)T] = \underline{K}^b(nT) + \dot{\underline{K}}^b(nT)$$

Objectives

Following conventional practice, we choose as the objective maximizing social welfare, defined as the expected net present value of net revenues from all participants, minus all capital costs. This is the expected value of total participants' profits, and also corresponds to expected long-run producers' plus consumers' surplus.^{A.6}

Until Section 3.5 we will continue to ignore transactions costs. Thus the objective function at time 0 is to maximize, over all relevant decision variables, and subject to constraints:

$$(3.A.10) \quad W = E_0 \sum_{n=0}^{\infty} e^{-rnT} \sum_{j \in \emptyset} [NR_j(nT/\underline{K}_j, \tilde{a}) - I_j] + NR^b - I^b$$

where $E_0 =$ expectation operator based on all information available at time 0.

This problem decomposes neatly, especially when full spot pricing is used. Define optimal short-term welfare for cycle n as:

$$(3.A.11) \quad \text{WST}(\underline{K}_j(nT), \underline{K}^b(nT); j \in \emptyset) \\ = E_{(n-1)T} \sum_{j \in \emptyset} NR_j(nT) + NR^b(nT)$$

subject to constraints (3.A.1a) to (3.A.8).

The global social welfare (3.A.10) decomposes to a series of short-run welfare measures which are conditioned on available capital stock plus investment decisions:

$$(3.A.12) \quad W = E_0 \sum_n e^{-rnT} \text{WST}(\underline{K}_j(nT), \underline{K}^b(nT), nT; j \in \emptyset) \\ - E_0 \sum_n e^{-rnT} \left[\sum_j I_j(\underline{K}_j(nT), \underline{K}_j(nT), nT) + I^b(\underline{K}^b, K^b) \right]$$

subject to

$$(3.A.4) \quad \underline{K}_j[(n+1)T] = \underline{K}_j(nT) + \dot{\underline{K}}_j(nT)$$

$$(3.A.9a) \quad \underline{K}^b[(n+1)T] = \underline{K}^b(nT) + \dot{\underline{K}}^b(nT)$$

Therefore, if decisions about appropriate operating variables (generator outputs and prices) can be postponed until the cycle to which they apply, the global social welfare maximization problem can be decomposed into a series of short-term problems, plus a master problem involving the choice of capital stock.

The short-term welfare maximization problem is crucial. Expanding (3.A.11) and appending the constraints gives the Lagrangian form:

$$(3.A.13) \quad \text{WST}(nT) = \text{Max } E_{(n-1)T}$$

$$\sum_j F_j(Y_j(t)/\underline{K}_j(nT), \underline{\bar{a}}(t); t = (n-1)T+1, \dots, nT) \\ - \sum_j R_j(r_j(t), Y_j(t)/\underline{K}_j, \underline{\bar{a}}(t); t = (n-1)T+1, \dots, nT)$$

$$\begin{aligned}
& + \sum_{t=(n-1)T+1}^{nT} \theta(t) \left[+ L(t) - \sum_j (1 - r_j(t)) Y_j(t) \right] \\
& - [\underline{Z}(t) - \underline{Z}_{\max}(t)] \underline{\mu}^Z(t) + [\underline{Z}(t) - \underline{Z}_{\min}(t)] \underline{\eta}^Z(t) \\
& - \sum_j [Y_j(t) - Y_{\max,j}(t)] \mu_j^Y(t) - [Y_j(t) - Y_{\min,j}(t)] \eta_j^Y(t) \\
& r_j(t) = 0 \text{ or } 1
\end{aligned}$$

which has duality conditions at each period t of:

$$(3.A.14) \quad \theta(t) [-L(t) + \sum_j [Y_j(t) - Y_j^R(t)]] = 0$$

$$\underline{\eta}^Z(t) [\underline{Z}(t) - \underline{Z}_{\min}(t)] = 0$$

$$\underline{\mu}^Z(t) [\underline{Z}_{\max}(t) - \underline{Z}(t)] = 0$$

$$\mu_j^Y(t) [Y_j(t) - Y_{j,\max}(t)] = 0$$

$$\eta_j^Y(t) [Y_{j,\min}(t) - Y_j(t)] = 0$$

where θ , $\underline{\mu}^Z$, $\underline{\eta}^Z$, μ_j^Y , $\eta_j^Y \geq 0$ are the shadow prices

(Lagrange multipliers) on the corresponding constraints. ^{A.7}

Solving the Model

I will now solve the model just presented. As the electric power system evolves in response to the exogenous stochastic processes $\tilde{a}(t)$

and $\tilde{b}(t)$, the welfare-maximizing utility responds by adjusting its various control variables. The more controls it has and the more information it has, the better it can do. Furthermore, the more often it can adjust the controls, the better it can do, since the effects of forecasting error and time aggregation will be reduced. For this section I will assume that most participants receive individual full spot prices which are set in real time by the central utility. I assume that the utility has direct control over T and D investments \underline{K}^b , but none over individual generation $Y_j(t)$ or over individual investments \underline{K}_j . Furthermore, I assume that the utility has no direct knowledge of participants' value added functions F_j nor of their capital stock \underline{K}_j , nor of the random process $\tilde{a}(t)$. It will turn out that the utility can control generators optimally using full spot prices, even if it does not own them. Thus the distinction between central and independent ownership is immaterial, as long as independent units acting as price takers are under full spot prices. (I will show what happens when either condition is violated.)

In this appendix I will look only at participants which are on on full spot pricing. Other participants will be modeled as contributing a "background demand" which is not responsive to the current spot price. The central controller can influence their behavior by rationing, and optimal rationing policies will be shown. Optimal pricing for participants not on full spot pricing was covered in Section 3.3.

Short-Term Maximization

As mentioned, welfare maximization decomposes into a series of short-run problems (equation 3.1.13) tied together by a long-run problem involving capital stocks (equation 3.1.12). I will solve this in steps. Here I derive the optimal full spot prices and rationing rule for each participant and show that proper prices lead to socially optimal behavior Y_j , without a need for rationing. Then I will discuss optimal rationing for participants not on full spot pricing. Finally I will discuss optimal investment.

At each moment, the central price setter wishes to maximize short-term social welfare cost as measured by equation (3.A.13). WST is a function of installed capital stocks, of the random variables $\tilde{a}(t)$ and $\tilde{b}(t)$ for the rest of the cycle, and the desired and rationed demand levels, $Y_j(t)$ and $Y_j^R(t)$ for the rest of the cycle. Some of these enter directly into the value added and rationing functions F_j and R_j ; the others affect the constraint equations. The $Y_j(t)$ cannot be controlled directly, but can be influenced by changing the spot prices, p_j , for the rest of the cycle, for those customers on spot pricing.

To find optimal $p_j(t)$, I will examine the first-order conditions for generation levels $Y_j(t)$. First-order conditions for private profit maximization and for social welfare maximization differ by several terms. By proper choice of spot prices these differences can be reduced to zero, giving convergence of profit and welfare maximizing behavior.

Consider participant j 's behavior $Y_j(s)$ at time s . Its objective is to maximize expected $NR_j(T)$ in equation (3.A.2) subject to constraint (3.A.5), conditional on all past values of $Y_j(t)$, $\underline{a}(t)$, and \underline{K}_j . The Lagrangian is:

$$(3.A.15) \quad E_s NR_j(T) - \mu_j^Y(s)[Y_j(s) - Y_{\max,j}(s)] \\ - \eta_j^Y(s)[Y_{\min,j}(s) - Y_j(s)]$$

The first-order condition for this is:

$$(3.A.16) \quad 0 = E_s \sum_{t=1}^T \left\{ \frac{\partial Y_j(t)}{\partial Y_j(s)} \frac{\partial F_j}{\partial Y_j(t)} + p_j(t) - \mu_j^Y(t) + \eta_j^Y(t) \right\}$$

where $\mu_j^Y(t)$ is the shadow price on the capacity bound $Y_{\max,j}(t)$ in (3.A.5) and $\eta_j^Y(t)$ is the shadow price on the capacity bound $Y_{\min,j}(t)$ in (3.A.5).^{A.8} The summation occurs because Y_j now influences optimal Y_j later.

Notice that $E_s \partial Y_j(t) / \partial Y_j(s)$ is 0 for $t < s$, 1 for $t = s$, and is a complex function of anticipated $p_j(t)$ and $\underline{a}(t)$ for $t > s$. That is, the participant must consider how its present action $Y_j(s)$ will affect its future optimal decisions. This can be a complex dynamic programming problem; see Chapter 4.

Equation (3.A.16) implies a "demand function" mapping $Y_j(s)$, $p_j(t) / \underline{K}_j$, $\underline{a}(t)$, $t = s, \dots, T$. This mapping may not be unique, but the capacity constraints ensure that at least one solution exists.^{A.9} Notice that $Y_j(s)$ is a demand functional, not a function, since it depends on the stochastic process $p_j(t)$, rather

than on the actual realization of prices. One consequence is that traditional long-run elasticities can be defined in many different ways. The most useful here is the response of current demand to changes in the parameters of the stochastic process $p_j(t)$, rather than to deterministic changes in a single price.

Now consider the social short-run maximization problem (3.A.13). It has first-order conditions:

$$(3.A.17) \quad 0 = \frac{\partial WST}{\partial Y_j(s)} = E_s \sum_{t=1}^T \frac{\partial Y_j(t)}{\partial Y_j(s)} \left\{ \begin{aligned} & \frac{\partial F_j}{\partial Y_j(t)} - \frac{\partial R_j}{\partial Y_j(t)} - \mu_j^Y(t) + \eta_j^Y(t) + \left[1 - \frac{\partial L}{\partial Y_j(t)} \right] \theta(t) \\ & + \frac{\partial Z(t)}{\partial Y_j(t)} [\underline{\eta}^Z(t) - \underline{\mu}^Z(t)] \end{aligned} \right\} \quad A.10$$

There is no way for a central controller to solve (3.A.17) directly since it does not know the functions F_j or the stochastic arguments $\tilde{a}(t)$. However, comparing (3.A.17) with (3.A.16) we see that the two problems are equivalent if the utility sets full spot prices:

$$(3.A.15) \quad p_j(s) = p_j^*(s) = \left[1 - \frac{\partial L(s)}{\partial Y_j(s)} \right] \theta(s) + \frac{\partial Z(s)}{\partial Y_j(s)} [\underline{\eta}^Z(s) - \underline{\mu}^Z(s)]$$

In effect participant j causes externalities on others via the system constraints on losses and \underline{Z} . Optimal full spot prices internalize all those effects through $p_j^*(s)$ in (3.A.18). This equation was discussed further in Section 3.2.

Actual Calculation of θ

How can a central utility actually calculate $p_j^*(t)$ in real time at each moment t ? There are two issues: calculating $\theta(t)$, which is common to all participants j , and calculating the participant-specific adjustments, i.e., those terms involving losses and voltage/line flow constraints. Incremental losses, $\partial L / \partial Y_j$, are already calculated by most utilities for major points in their T and D network. The terms stemming from constraint (3.A.7), however, require some estimate of local behavior of individual demands, $\partial Y_j / \partial p_j$. Fortunately, these terms are only rarely positive on most systems. Exact calculation of these may be difficult or impossible for some participants. But as the voltage and line flow constraints are somewhat elastic for short periods, exact calculations are not essential. Trial and error can be used. As usual, the utility will be better able to avoid violating the constraints (3.A.7) under participant-specific full spot prices, than under any other pricing method.^{A.11}

The $\theta(t)$ term will dominate spot prices and is most important. It can be calculated in two ways, one corresponding to the operation of a competitive market, and the other to current utility practice. The competitive approach is to think of the utility as an auctioneer which tries to find $\theta(t)$ such that supply and demand balance, i.e., constraint (3.A.8) is satisfied. In a well-behaved utility system there will be a unique solution. The role of a real-time auctioneer in a non-Walrasian market is complex (Grossman, 1981). Nonetheless, such markets do clear every day. The more information the

utility/auctioneer has about F_j for j on spot prices, the better it will do.

The other approach to $\theta(t)$ is based on the concept of a marginal generating plant. Consider a utility with all generators on full spot pricing, but other participants on predetermined prices. Let m be the index of a marginal generator. That is, $Y_{\min,m}(t) < Y_m(t)$ so $\eta_m^Y(t) = 0$. Also

$$(3.A.19) \quad Y_m(t) < Y_{\max,m}(t) \text{ and } \mu_m^Y(t) = 0$$

or all capacity is in use:

$$(3.A.20) \quad Y_j(t) = Y_{\max,j}(\tilde{t}/\underline{a}(t)) \quad \forall j \in \phi^+$$

$$\mu_m^Y(t) > 0$$

Define the system λ as:

$$(3.A.21) \quad \lambda(t) = \frac{E_t \frac{\partial F_m}{\partial Y_m(t)} + (\underline{\eta}^Z - \underline{\mu}^Z) \frac{\partial Z(t)}{\partial Y_m(t)}}{\frac{\partial L(t)}{\partial Y_m(t)} - 1}$$

Define the "rationing premium" as:

$$(3.A.22) \quad \mu(t) = \theta(t) - \lambda(t)$$

Substituting $\theta = \lambda + \mu$ in (3.A.18) gives

$$(3.A.23) \quad p_m^*(t) = - E_t \frac{\partial F_m}{\partial Y_m(t)} + \mu(t) \left[1 - \frac{\partial L(t)}{\partial Y_m(s)} \right]$$

But we know from the profit maximization behavior (eq. 3.A.16) that

for the current period, t , the generator m will act so that:

$$(3.A.24) \quad p_m(t) = - E_t \frac{\partial F_m}{\partial Y_m(t)} + \mu_m(t) - \eta_m(t)$$

(Recall that F_j is negative for a generator.)

Hence we can define the key determinants of spot prices as:

$$\theta(t) = \lambda(t) + \mu(t) = \text{social value of another kwh}$$

$$\lambda(t) = \text{system } \lambda \text{ (equation 3.A.21)}$$

= marginal operating cost of last unit loaded, adjusted for line losses and constraints.

$$(3.A.25) \quad \mu(t) = \text{total generating capacity shadow price} = \begin{cases} 0 & \text{if } Y_j(t) < Y_{j,\max}(t, K_j, \tilde{a}_j) \\ & \text{for some } j \in \phi^+ \\ \mu_m(t) / [1 - \frac{\partial L(t)}{\partial Y_m(t)}] & \text{if all} \end{cases}$$

capacity is in use.

The optimal spot price formula (3.A.15) becomes:

$$(3.A.25b) \quad p_j^*(s) = [1 - \frac{\partial L(s)}{\partial Y_j(s)}] [\lambda(s) + \mu(s)] + \frac{\partial Z(s)}{\partial Y_j(s)} [\underline{n}^Z(s) - \underline{\mu}^Z(s)]$$

The full spot price to j is the system λ , adjusted for j 's impact on system losses and on system line constraints (which may be positive or negative, depending on j 's "location" in the T and D network and the state of the T and D system), plus a term to curtail demand if total system capacity is fully used.^{A.12}

The difference in price between two points at one time was discussed extensively in Sections 3.1 and 3.2.

Rationing

Optimal full spot prices $p_j^*(t)$ can be calculated in real time for each participant, at least approximately. However, because of transactions costs it will not be desirable to have all participants on full spot pricing. These participants may have to be rationed, if their unrationed demand/generation levels would lead to violation of constraints on line flows, voltages, or energy balance.

I assume the central utility can ration participants only by opening a circuit breaker, i.e., all or nothing. Suppose that participant i sees non-spot prices $p_i(t)$ (see Section 3.3). Substituting (3.A.15) into the short-term welfare equation (3.A.13) gives the difference in welfare if i is rationed ($r_i = 1$) of:

$$\begin{aligned}
 (3.A.26) \quad & WST(r_i(t) = 0) - WST(r_i(t) = 1) \\
 &= p_i^* Y_i(t) + ER_i(r_i(t) = 1/Y_i(t)) \\
 &\quad - ER_i(r_i(t) = 0/Y_i(t)) \\
 &\quad - \mu_i^Y(t)[Y_i(t) - Y_{\max,i}(t)] \\
 &\quad + \eta_i^Y(t)[Y_i(t) - Y_{\min,i}(t)]
 \end{aligned}$$

where

$Y_i(t) = Y_i(p_i(t), t)$ = demands chosen by i at prices

$p_i(1), \dots, p_i(T)$

$R_i(r_i(t)/Y_i(t))$ = disruption and loss of value added

caused by $r_i(t)$, holding rationing in all other periods fixed, when demand would have been $Y_i(t)$.

If this expression is negative, then rationing of i is optimal in period t . For simplicity, consider the case of rationing a net user during t , with no rationing needed at other periods of the cycle, and with no binding $Y_{\min,i}$ constraint. Then $Y_i(t) < 0$, and

$$\begin{aligned} \text{WST}(0) - \text{WST}(1) < 0 & \quad \text{IFF} \\ (3.A.27) - p_i^*(t)Y_i(t) > R_i[r_i(t) = 1/Y_i(p_i(t))] = R_i(1) \end{aligned}$$

Social value of electricity used $>$ Disruption caused by rationing. This inequality is more likely to be satisfied, the greater $p_i^*(t) - p_i(t)$. Since self-curtailment to 0 is always a possibility, it must be true that:

$$(3.A.28) \text{ Cost to } i \text{ of electricity used} = - p_i(t)Y_i(t) < R_i(1)$$

Therefore when $p_i^*(t) \leq p_i(t)$, (3.A.25) cannot hold; rationing is not optimal.

Consider a fixed price $p_i(t)$ and resulting $R_i(1)$. Suppose total system demand rises or generating capacity falls. As $\tilde{a}(t)$ and $\tilde{b}(t)$ lead to higher $\theta(t)$ and therefore higher $p_i^*(t)$, eventually $p_i^*(t)Y_i(t) + R_i(1)$ will change sign; rationing will become optimal. A.13

In practice the utility will have only a rough idea of $R_i(1)$, and it will not be able to control $r_i(t)$ individually for small participants. Therefore once $\theta(t)$ reaches some threshold corresponding to the estimated "average disruption" caused by rationing across a group of participants, it will be optimal to stop

raising $\theta(t)$, hence hold steady the price to spot participants, and instead ration the non-spot participants.

Thus under optimal utility behavior:

- o The possibility of rationing puts an upper bound on spot prices.
- o The more participants are on spot pricing, the less often rationing will be needed for other participants.
- o The probability that i will be rationed is an increasing function of $p_i^*(t) - p_i(t)$. In particular, multiple rate classes may exist with different rules for updating their prices (see Section 3.5). Then $p_i^*(t) - p_i(t)$ will be largest for the rate classes whose price is updated least often. Therefore, all else equal, participants on infrequently updated prices will be rationed most often.

Investment in Generation

Consider the level of the optimal capital stocks, \underline{K}_j and \underline{K}^b . These are found by "solving" the long-run welfare maximization problem (3.1.12). Again there are two routes to achieving this capital stock: direct investment in generating facilities by a welfare-maximizing central utility, or using full spot prices for sales to and from decentralized participants. Either approach yields the same result, in the absence of:

- o Different forecasts by the central utility and participants.
- o Prices which differ from optimal full spot prices.

- o Scale economies in plant level capital costs, plus market power of participants.

The optimal level of \underline{K}_j during cycle n is found by differentiating (3.A.12) with respect to investment $\dot{\underline{K}}_j$ during cycle $n-1$. The resulting first-order conditions are complicated by the dynamic and stochastic nature of the problem. Nonetheless the optimal investment rule reduces to the familiar "invest up to the point that the expected marginal cost of investment equals the expected marginal benefit, both properly discounted." The expected marginal benefit of investment is an improvement in the short-term welfare possible for each realization of \tilde{a} and \tilde{b} , i.e., an improvement in the (stochastic) short-term production-possibility frontier.

The optimal level of investment in generation has been discussed by previous authors, as discussed in Section 3.4. The use of spot pricing does not alter fundamental optimality conditions. However the use of spot prices leads to a new and much more intuitive interpretation. Basically, a generator should be built iff it will have an expected positive profit when paid full spot prices. Also, investment in storage and end use capital is covered by the same equations.

Substituting in (3.A.13) and differentiating (3.A.12) with respect to \underline{K}_j to find socially optimal investment in cycle 0 gives first-order conditions:

$$(3.A.29) \quad \frac{\partial I_j(\dot{\underline{K}}_j(0), \underline{K}_j(0), 0)}{\partial \dot{\underline{K}}_j(0)} + E \sum_{n=1}^{\infty} e^{-rnT} \frac{\partial I_j(nT)}{\partial \dot{\underline{K}}_j(0)}$$

$$= E \sum_{n=1}^{\infty} e^{-rnT} \frac{\frac{\partial K_j}{\partial K_j}(nT)}{\frac{\partial K_j}{\partial K_j}(0)} \frac{\partial}{\partial K_j(nT)} WST(K_j(nT), \underline{K}^b(nT), nT; j \in \emptyset)$$

The first term is the marginal cost of the investment now. The second set of terms on the left-hand side is the discounted marginal impact of investment now on investment costs later: learning curve effects, using up hydro sites, etc. Properly assessing the left-hand side is important in practice, but its theoretical interpretation is well-known and it won't be discussed further.

Each term of the summation on the right-hand side is the expected net present value of short-term welfare improvements due to increased investment now. $E \frac{\partial K_j}{\partial K_j}(nT) / \frac{\partial K_j}{\partial K_j}(0)$ reflects capital stock depreciation and the effects on anticipated future investment on investment now. $\frac{\partial WST}{\partial K_j}$ is the improvement in realized short-term welfare due to a larger capital stock. Both factors are stochastic, and their evaluation requires forecasts about the future, demand growth, fuel prices, and other elements of $\tilde{a}(t)$ and $\tilde{b}(t)$.

Differentiating (3.A.13) gives for $n \geq 1$:

$$(3.A.30) \quad \frac{\partial WST}{\partial K_j} = E \frac{\partial F_j}{\partial K_j} - \frac{\partial R_j}{\partial K_j} + \sum_{t=1}^T \frac{\partial Y_{\max,j}(K_j, \tilde{a}(t))}{\partial K_j} \mu_j^Y(t) \\ - E \sum_t \frac{\partial Y_{\min,j}(K_j, \tilde{a}(t))}{\partial K_j} \eta_j^Y(t) \quad \text{A.14}$$

The first two terms are the expected value of increases in value added (decreases in non-electrical costs or increases in final production) and reductions of losses due to involuntary rationing.

For example, a generator may invest in a long-term coal contract or in discretionary maintenance. A user may invest in conservation equipment. A.15

The third term is especially important for capacity expansion by a generator. From (3.A.16) and (3.A.17) we have that $\mu_j^Y(t)$ is the instantaneous value of another unit of generation at t :

$$(3.A.31) \quad \mu_j^Y(t) = \begin{cases} 0 & \text{if } Y_j(t) < Y_{j,\max}(t) \\ p_j^*(t) + \frac{\partial F_j}{\partial Y_j(t)} - \frac{\partial R_j}{\partial Y_j} & \text{otherwise} \end{cases}$$

This is marginal net revenue. It is the value of the option to produce electricity at marginal cost $\partial F_j / \partial Y_j$ and sell it at $p_j^*(t)$.

Increasing K_j permits j to have $\partial Y_{\max,j}(t) / \partial K_j$ of such options for each unit of capacity. This derivative is proportional to unit availability, $\tilde{a}_j(t)$.

The shaded portion of Figure 3.1.3 shows the value of this option, for a kilowatt of generating capacity which is always available. The curved line is the full spot price at each moment, conditional on unit j being available. The value of a stochastically available kw of capacity is the shaded area integrated only over t such that $\tilde{a}_j(t) = 1$. Thus the value of plants of different reliability can be quantified. A.16

Figure 3.1.3, which shows p_j as a function of time, can be translated into a Price duration curve, just as a load curve can be translated into a load duration curve. Such a curve was shown in Figure 3.1.4. The area above $\partial F_j / \partial Y_j$ in the price duration curve

then measures $E\mu_j^Y(t)$, conditional on the rest of the capital stock. This provides a quick means of approximating $\partial WST/\partial K_j$ for a new generator [Caramanis, 1981].^{A.17}

Profit Maximizing Investment Behavior

Profit maximizing investment behavior will equal social welfare maximizing investment if the participant has the same anticipated price probability distribution, is on full spot pricing, and has no market power. To see this, consider the profit maximand (3.A.3). Profit maximization of (3.A.3) with respect to K_j gives first-order conditions identical to (3.A.27), except that $\partial WST/\partial K_j$ is replaced by $\partial NR_j/\partial K_j$. These two appear identical except for an additional term

$$E \sum_t \frac{\partial p_j(t)}{\partial K_j} Y_j \leq 0.$$

That is, profit maximizing participants with market power will consider that their investments may affect the price. If investments are continuously variable in size, the $\partial p_j/\partial K_j$ term may be inconsequential even if Y_j is large enough to affect prices. The reason is that investment by j may preempt subsequent investment by its competitors. (However, if j is a regulated monopoly, entry may be prohibited and $\partial p_j/\partial K_j$ may be large enough to motivate rampant underinvestment. See Section 3.6.)

But if investments are lumpy due to indivisibilities or

plant-level economies of scale, an additional wedge exists between private and social investment criteria. Suppose j builds a unit of size \mathring{K}_j . Then for a typical realization of $\tilde{a}(t)$, $\tilde{b}(t)$, this will shift out the industry supply curve as shown in Figure 3.A.2. If the new unit has constant marginal operation cost $\partial F_j / \partial Y_j$ as shown, then $NR_j(t)$ is the area A. But the change in WST during period t is $A + B$. Note that the size of B is quadratic in \mathring{K}_j , so this effect depends critically on the lumpiness of optimal investments. The effect is to reduce investment below the socially desirable level.

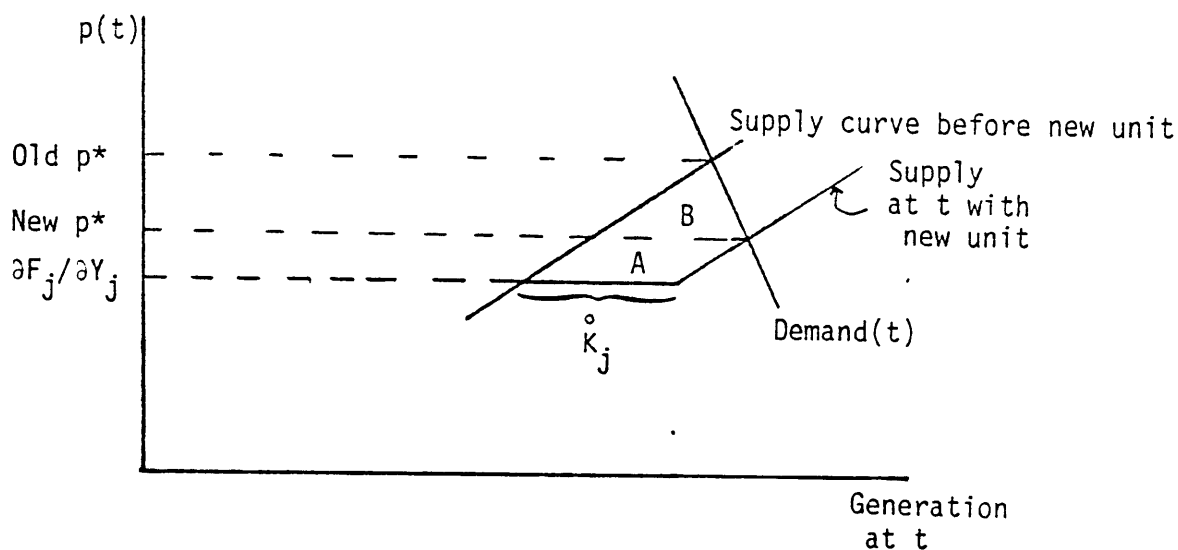


Figure 3.A.2

Private vs. Social Investment Criteria

Furthermore, profit maximizing firms may evaluate (3.A.30) at different levels of Y_j and p_j , leading to very different private and social values for the same investment. One case is if $p_j(t) \neq$

$p_j^*(t)$, i.e., j is not on full spot prices. Then Y_j , μ_j^Y , and η_j^Y may all be affected. The other case is a firm with short-run market power, which will have a lower generation/demand level Y_j at any $p_j(t)$ than if it were a pure competitor.

Thus, market power may lead to underinvestment. Prices other than full spot prices may lead to over- or underinvestment, depending on the sign of $p_j^*(t) - p_j(t)$. For realistic cases, it will lead to underinvestment.

Investment in Transmission and Distribution

Repeating the above procedures, this time with respect to investment \underline{K}^b , in strengthening the TD network, we get first-order conditions:^{A.18}

$$\begin{aligned}
 (3.A.32) \quad & \frac{\partial I^b}{\partial K^b(0)} + \sum_{n=1}^{\infty} e^{-rnT} \frac{\partial I^b(nT)}{\partial K^b(0)} \\
 & = E_0 \sum_n e^{-rnT} \frac{\partial K^b(nT)}{\partial K^b(0)} \sum_t -\theta(t) \frac{\partial L(t)}{\partial K^b} \\
 & \quad + \frac{\partial Z_{\max}(t)}{\partial K^b} [\underline{n}Z(t) - \underline{\mu}Z(t)]
 \end{aligned}$$

Thus the value of the investment is the expected discounted value of reduced losses it causes plus the value of reduced TD system stringency conditions. Naturally, these will be larger, the weaker the portion of the system which is augmented, or the more heavily that portion is used. Thus the larger the price difference between the

points connection by a new transmission line (see Section 3.2), the more valuable the line.

Other Control Methods: Comment

Two standard alternatives to full spot pricing are time-of-use (predetermined) pricing, and rationing. As long as transactions costs of spot pricing are zero, both approaches are dominated.

In theory, equally good short-run behavior could be achieved by rationing instead of spot pricing. However, the efficient information "collection" provided by price mechanisms leads to spot pricing dominating quantity allocation techniques. In order to achieve socially optimal short-run behavior by means of rationing, the central utility would have to know the full function $F_j(Y_j(t)/K_j, \tilde{a})$ conditional on the current values of the last two arguments. This is feasible for large electricity producers; hence utilities own and dispatch (set Y_j) for such generators. But it is not realistic for thousands of small electricity users and producer. With j on spot pricing, the central utility has to know at most the current and local behavior of $dY_j(t)/dp_j(t)$. This is much less information, and can be found roughly by trial and error.^{A.19}

Investment behavior is more problematic than short-run behavior, because it requires longer-range information. To the extent that the central utility makes available its own information and forecasts about future full spot prices, and to the extent that participants are profit-maximizers with rational expectations, then full spot pricing

with decentralized investment decisions, will give results as good as centralized investment decisions, assuming perfect competition. Again, for any information structure, full spot pricing is at least as good as rationing and in practice will be better.^{A.20} But a social welfare maximizing central utility may make investment decisions differently than a profit-maximizing firm with significant market power. When such divergence occurs, it will always be in the direction of underinvestment by the profit-maximizer.

Of course, full spot pricing dominates predetermined prices, except for their relative transactions costs. Adding more frequent price changes must reduce the time aggregation error. Both short-run and long-run (investment) behavior cannot be socially less desirable under full spot pricing.^{A.21} Chapter 5 will quantify this.

All of this is a recapitulation of Pareto optimality of properly set prices, under conditions of dynamic market evolution and uncertainty. But note that the same comparisons apply to the control method used for individual participants. And the same equations apply under systems which mix full spot prices and predetermined prices.

FOOTNOTES

A.1 Actually many participants switch back and forth from net generation to net use. The notation used will cover this case.

A.2 Panzar [1976] uses a similar neoclassical production function, but for total generation rather than individual participants. We assume that F_j is continuously differentiable and locally strictly concave.

A.3 This assumption corresponds to the use of a rotating blackout. Customers with two circuits, one interruptible, may be loosely modeled as two independent customers, one of which gets interrupted.

A.4 I will ignore discounting within cycles, since a cycle will usually be interpreted as a month or less.

A.5 Both $L(t)$ and $Z(t)$ also depend on reactive flows in the network. See Caramanis, Bohn, Schweppe [1982]. A reference on losses is Elgerd [1977 or 1982].

A.6 See footnote 1.5 of Chapter 3.

A.7 Equation (3.A.14) is in "standard form," so that when constraints $g(Y) \leq 0$ are taut, the corresponding shadow prices will be positive. Caramanis, Bohn, Schweppe [1982] did not rigidly follow this convention; therefore its equations will differ in signs.

A.8 In this form the dual variables μ_j^Y and η_j^Y are non-negative, with the usual complementary slackness condition.

A.9 For participants which are generators, the technology guarantees an upward sloping marginal heat rate and therefore marginal generating cost $-\partial F_j / \partial Y_j$ as Y_j approaches $Y_{\max,j}$. See for example El-Hawary and Christensen [1979]. Therefore there will be a unique maximum of (3.A.15). Furthermore, changes in $p_j(s)$ will produce continuous shifts in $Y_j(s)$.

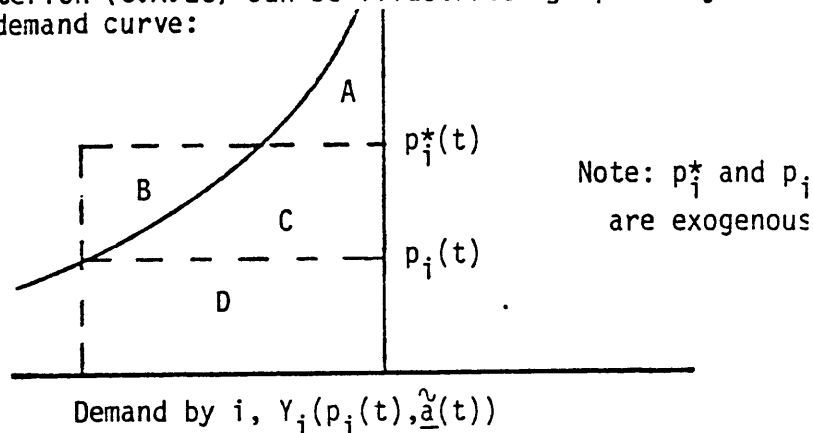
The same is not necessarily true for individual electricity consumers, as shown in Chapter 4. But aggregate demand will generally be a smooth function of price, ensuring the existence of a solution to the optimal spot pricing equation developed below.

A.10 This is conditional on $r_j(s) = 0$, i.e., no rationing. If $r_j(s) = 1$, then the price $p_j(s)$ is irrelevant.

A.11 Here, however, quantity based controls may in theory be better than price based. This is the point of Weitzman [1974]. He assumes that controls (prices or quantities) must be chosen before all uncertainty is resolved, whereas in this dissertation timing is a decision variable, and decisions can be postponed until most or all information is revealed. To the extent that large irreducible uncertainty about participant response still exists under full spot prices, rationing or other quantity control may be superior. Such uncertainty will be larger (in percentage terms) for individual participants than for aggregates of participants. The use of quantity controls such as microshedding [Schweppe et al., 1979] requires further research.

A.12 If the utility knows F_j for $j \in \phi^+$, which it will know if it owns the generators, it can determine which generator is marginal given the current level of demand, and from (3.A.18) find λ . This is essentially what utilities do today. Thus, under normal conditions, i.e., when $\mu = 0$, a conventional integrated utility already calculates the spot prices, under the assumption that demands are unresponsive to spot price. With only a few net users on full spot pricing, this is a good approximation as long as $d\lambda(t)/dt$ is small. As more net users go on full spot prices, accurate forecasting of demand responsiveness to spot prices becomes more important.

A.13 Criterion (3.A.28) can be illustrated graphically for a one-period demand curve:



The social value of electricity used is areas $B + C + D$ (price + quantity). The cost of rationing is at least $(A + C + D) = F_i(Y_i)$ (by equation 3.A.1). Hence, if $B < A$, rationing is not optimal. Clearly, the smaller $p_i^* - p_i$, the more likely that $B < A$, and the less desirable is rationing.

A.14(3.A.30) is a term-by-term differentiation of (3.A.13), using the envelope theorem to ignore $\partial Y_j / \partial K_j$.

A.15These first-order conditions are valid for continuous investment with a constant or increasing marginal cost of investment. Otherwise discrete optimization methods (e.g., enumeration) must be used in place of (3.A.29). In practice elaborate integer programming methods can be used, as in Scherer [1976, 1977]. All of the discussion still applies to the discrete case.

A.16The last term of (3.A.30) measures the value of the option to reduce Y_j . For example, if j is a pumped hydro unit, increasing the size of the turbine/pump will increase the value of electricity used during the pumping phase. For a conventional generator, this term depends on turndown constraints.

A.17Greater uncertainty about future fuel prices and demands will affect the shape of the price duration curve, which should therefore not be based on a single "most likely" forecast. See Ellis [1981].

A.18Using $Z_{\min}(t) = Z_{\max}(t)$.

A.19Even this local responsiveness need be known mainly when one of the system capacity constraints is binding. The rest of the time, it is adequate for the utility to know total price responsiveness of all participants: $d \sum_j Y_j / dp(t)$.

j

A.20Grossman [1981] goes even farther to show that with rational expectations plus complete contingent claims markets, the results of full spot pricing are as good as those achievable by a planner with all of the economy's information. However, the assumption of complete contingent claims markets is too strong for any realistic organization of electricity markets. If a central utility "hoards" its own information, it might be able to make better long-range decisions than anyone else. A welfare maximizing central utility would therefore reveal its plans, market surveys, and other information it has collected to anyone willing to pay the cost of searching its files.

A.21In pathological cases a given investment could produce socially better effects under non-spot prices. See footnote 3.7 of Chapter 5. But in Chapter 4 I argue that spot pricing would still make investment more likely.

CHAPTER FOUR

OPTIMAL BEHAVIOR UNDER SPOT PRICES

The value of different time-varying rates depends on how much they affect customer behavior. If a customer will not respond to change in spot prices, there is no social value to putting it on spot pricing. This chapter analyzes customer behavior under spot and other time-varying prices. Several useful theoretical concepts emerge. In addition, this chapter forms the basis of the case study simulations of Chapter 5.

Contributions of This Chapter

Specific contributions of the chapter include the following.

- o It develops a rigorous microeconomic model of firm behavior under stochastic factor prices. From the model, it is possible to calculate the value to different firms of prices generated by different stochastic processes. Two polar forms of response to prices are derived.¹ The model is used to predict what kinds of customers and utilities will show the largest effects from spot pricing. Although I do not do so here, the model can also be used to analyze behavior under quantity-oriented load management techniques. (See Section 3.4.)

- o Conventional wisdom says that the storage of electricity is feasible only in expensive, centralized facilities. This chapter argues that this is incorrect. Customer behavior that is functionally identical to decentralized storage of electricity is technically

possible and sometimes inexpensive, and will occur given incentives in the form of time-varying prices. Central storage can have a large effect on optimal generating capital stocks and costs [Nguyen, 1976], and the same is true for decentralized storage.

- o The model developed here has two applications other than deriving theoretical results. First, it is used in Chapter 5 to predict the behavior of individual hypothetical customers under various rates. Second, it can be used by actual firms which are deciding how to respond to time-varying prices.²

- o The chapter develops (but does not rigorously prove) several results about the incentives for investments under different rates. Spot pricing increases the value to both customers and utilities of certain customer investments.

- o Finally, there is skepticism among non-economists that customers would respond to prices which change stochastically every hour. Other groups have accepted that customers will respond to prices but have only analyzed behavior under deterministic conditions. This chapter cites examples of customer response to stochastic prices, and argues that if spot prices are broadly used, there will be large response from certain types of customers.

- o The chapter's analysis can also be applied to customer behavior under time-varying prices for other goods, such as natural gas and long-distance, one-way data communication. For example, decentralized customer "storage" of both goods is possible.

Alternate Approaches

Quite a bit of work has been done on estimating demand under deterministic time-of-use prices. Almost all of it is oriented toward econometric estimation of flexible functional forms, and toward the residential sector. The econometric approach to industrial demand behavior has the disadvantage that relevant historical data are required on enough customers to give statistically useful information. Because industrial processes vary widely both within and between industries, the only adequate data for long run industrial response estimates are probably European, which introduces complications.³ More important, only very special rates will give information about relative response to spot versus predetermined time-varying prices. Deliberate experiments using spot prices are still in the future and will be quite expensive. Section 4.4 discusses the existing rates which have elements of spot pricing. None of them are close to full spot prices.^{4,5}

An alternate approach to predicting response to time-varying prices is case-by-case analysis of real or hypothetical industrial plants. For simple time-of-use rates, profit-maximizing behavior modes can be determined essentially by inspection. This approach was used for several plants in seven different industries in Gordian Associates [1980], and also by Manichaikul and Schweppe [1980]. This approach has the advantage that it can be applied before any experimental evidence is available. It can also lead to the recognition of behavioral or other constraints on firm behavior which

were not obvious. However it does not necessarily lead to broadly applicable insights. It is also skilled-labor-intensive. Finally, it is possible to overlook optimal behavior modes when solving by inspection.

In this chapter I follow a third approach, based on mathematical programming models developed in the operations research literature for production planning decisions. These models take as exogenous the capital stock of a plant, other constraints on operations, the costs of various inputs, and the demand for final output, and solve for optimal production plans. Typically all costs are assumed constant over the planning horizon, and the minimal time period considered is one week, making these models not directly applicable. But with some effort, the same mathematical structures can be used to solve for optimal behavior as a function of time-varying prices. The resulting models can be used to predict the implications of different price variation patterns and types of production processes. They can also be used as algorithms for process control computers to optimize response to spot prices, in real time.

Section 4.1 presents a multi-period model of a general production process which has substantial short run flexibility despite a fixed capital stock. Section 4.2 shows how a variety of industrial and other electrical loads are covered by the model. Section 4.3 shows how to solve for optimal behavior and discusses the properties of the solution. Stochastic and predictable variations in electricity prices leave unchanged or improve the expected profits of the firms. Section

4.4 discusses experience with stochastic prices. Section 4.5 concludes by showing how the widespread application of full spot prices will tend to stabilize those prices.

4.1 Basic Model

The basic model is that of a multi-stage production process in a single plant, as shown in Figure 4.1.1. Suppose that the process has N stages connected in series as shown.^{1.1}

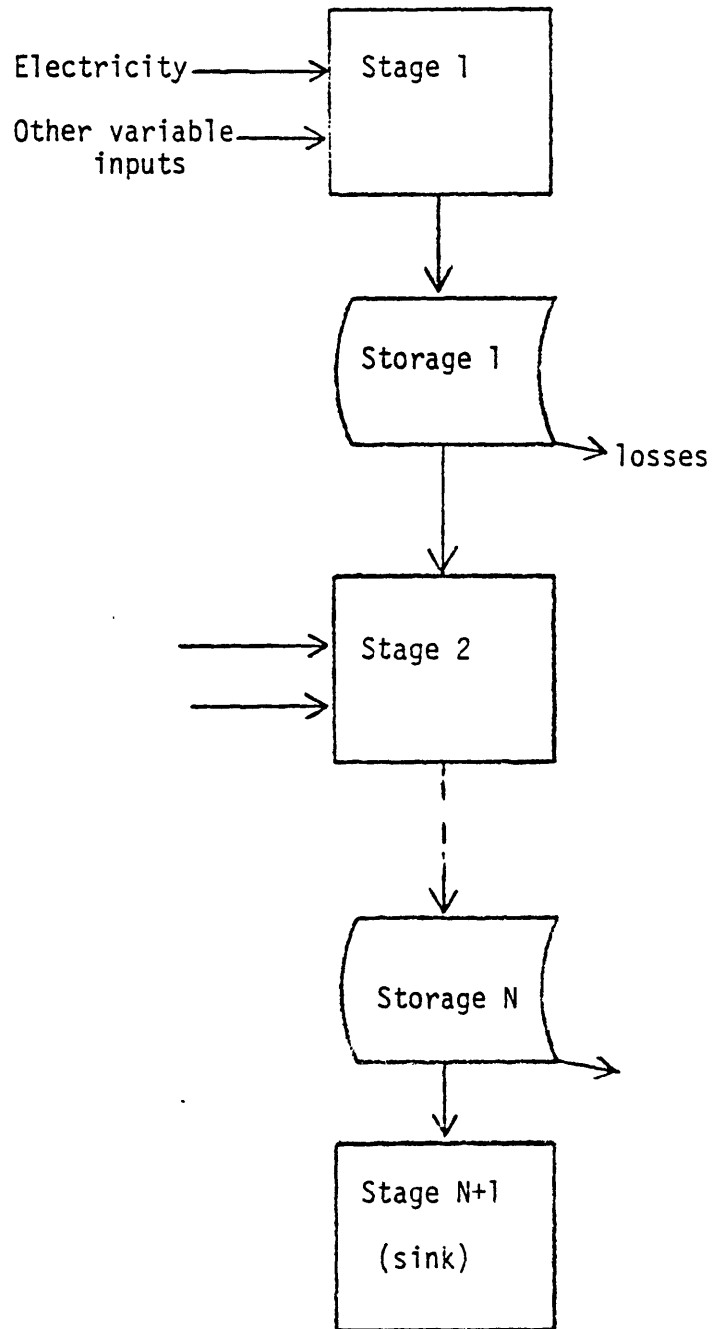
At each stage various inputs, including electricity and the output of the previous stage, are transformed into an intermediate product. The intermediate product can be stored in a storage buffer, perhaps with leakage. The output of stage N is a final product which is absorbed by the infinite sink, stage $N+1$.

Examples of processes which can be described in this way include air conditioning, electric arc furnaces, other thermal processes, electricity generation, fluids pumping, chemical processing, and many others. Specific examples will be discussed in Section 4.2. Obviously the same framework can be used to study the use over time of other inputs besides electricity.

The firm's basic optimal strategy is to turn the successive stages on and off as the prices of electricity and other inputs change, so as to reduce total production costs. Various constraints on maximum and minimum storage levels and on output from each stage are determined by the underlying capital stock of the plant, and determine optimal behavior. For example, if all storage buffers can hold only negligible amounts, then the operating modes are very limited: either operate all stages at once, or shut them all down.

Associated with each production stage are the costs of non-electric inputs and the amount of electricity used, per unit

Figure 4.1.1
Multi-Stage Production Process



processed.^{1.2} The final stage will have a negative "cost" per unit corresponding to the value of the final product. In this way, the firm can be modeled as a cost-minimizer, or a profit-maximizer with variable final production level.

An additional complication is that each stage may have several alternate methods, all of which use the same raw materials and produce the same intermediate product, but which have different efficiencies and capacities. For example, a pumping station may have several pumps; an air conditioning system may have several compressors. I will model this with a convex cost function.^{1.3}

Formal Statement of Problem

Define

- $p(t)$ = Electricity price to the firm, period t . May be spot or predetermined.
- $X_n(t)$ = Level of output of stage n during period t .
- $S_n(t)$ = Amount in storage buffer n at the end of period t .
- $e_n(X_n)$ = Electricity used in stage n to produce amount X_n in period t . A linear or convex function of the stage's output, up to the limit $X_{\max,n}$.
- $C_n(t)$ = Non electric variable costs per unit of production.
- L_n = 1- Loss coefficient for storage in buffer n .

Less than 1 when the intermediate "product" is some form of thermal energy.^{1.4}

- $X_{\min, N+1}(t)$ = Minimum final production each period. May be required to meet firm orders.
- $X_{\max, n}$ = Maximum production by stage n . These constraints are directly determined by installed capital stock for the plant. (K in Chapter 3.)
- $S_{\max, n}$ = Maximum storage capacity for buffer n . I will assume the minimum for each buffer is 0; no backorders.
- T = Planning horizon

With these definitions, the firm's production problem is as follows:

$$(4.1.1) \quad \text{Min } E \sum_{t=1}^T \sum_{n=1}^N [X_n(t)C_n(t) + p(t) e_n(X_n(t))] + X_{N+1}(t) C_{N+1}(t)$$

non-elec. costs
elec. cost
value of final product

subject to:

$$(4.1.2a) \quad S_n(t) = L_n S_n(t-1) + X_n(t) - X_{n+1}(t) \quad \text{Inventory balance}^{1.5}$$

$$(b) \quad X_n(t) \leq X_{\max, n} \quad \text{Production capacity}$$

$$(c) \quad 0 \leq X_n(t) \quad \text{Irreversible production}$$

$$(d) \quad X_{\min, N+1}(t) \leq X_{N+1}(t) \quad \text{Minimum final production}$$

- (e) $S_n(t) \leq S_{\max,n}$ Storage capacity
- (f) $0 \leq S_n(t)$ No backorders
- (g) Initial storage levels $S_n(0)$ given
for all $n = 1, \dots, N$
 $t = 1, \dots, T$ ^{1.6}

The problem is couched as a cost-minimization problem. Of course if the prices $p(t)$ are the optimal full spot prices for the firm, this coincides with social welfare maximization.

The firm's solution to this problem depends on what information it has about future $p(t)$. I will discuss this in Section 4.3.

Problems of this sort are common in the operations research literature. But usually $p(t)$ and $C_n(t)$ are assumed fixed over time, while final demand to be met, $X_{N+1}(t)$, varies. Inventories are then used to smooth production. Such models are surveyed in Johnson and Montgomery [1974, esp. Chapter 4]. It is rare to solve them as a function of input prices. Tsitsiklas [1979] has several useful mathematical results about this problem with time-varying prices, primarily for deterministic versions. As far as I know no one has discussed the interpretation or economic significance of these models with time varying prices, except Bohn [1981].

The problem (4.1.1, 4.1.2) is a dynamic programming problem. It is also a network flow problem. Each storage buffer at each period is a node, starting with a fictitious, infinitely full source node at time 1. Each node is connected to one or more arcs corresponding to

the methods usable at that stage.^{1.7} This allows the use of special methods for solving network problems. Such methods can be quite efficient. The stochastic nature of the problem increases the cost of solution, but special features of spot prices can be exploited to give close-to-optimal solutions cheaply.

4.2 Examples

This is an extremely general model. Virtually any firm can be modeled as a series of processing steps, some of which use electricity. A model on one time scale and level of detail can be embedded in a broader, less-detailed model. For example, any manufacturing plant can be modeled as a two-stage process, with the buffer between the stages corresponding to finished goods inventory, and the final stage consisting of goods shipped.^{2.1} The model's optimal solution would be to shut down the manufacturing process whenever the cost of variable inputs rises above the value of final production. Obviously, for many process which use relatively little electricity, only a very high spot price would lead to shutting down. Nonetheless, modeling the same firm in more detail might reveal subprocesses which would respond to the spot price.

I will discuss several examples in more detail. Several of these are used as the basis for the case studies of Chapter 5.

Example 1: Finished Goods Storage. Consider any process which produces a storable final product. Model it as a two-stage production process, with finished goods inventory the buffer between production and sales. A critical parameter is $s_1 = S_{\max,1}/X_{\max,2}$. It is the maximum number of periods for which production can be shut down without having to reduce sales. $C_2(t)$ is the price at which final goods are sold.

Examples of processes which fit this model include any

manufacturing process, as discussed above. Some particularly interesting ones are:

- o Pumped hydro generating plants. Here e_1' is the reciprocal of system efficiency. $S_{\max,1}$ is the reservoir size, measured in kWh of electrical energy output.
- o Gas liquefaction plants, cement plants, and others with a high ratio of electricity use per dollar of value added.
- o Municipal water systems. Here final demand $X_{\min,2}$ is stochastic and uncontrollable. C_2 is the social cost of not having water available when demanded. Water is pumped into holding tanks; final demand is by gravity feed. $S_{\max,1}$ is the capacity of municipal water tanks in gallons. $X_{\max,1}$ is the total pumping capacity in gallons per period. $e_1(X_1)$ is the efficiency of pumping in kWh per gallon. If several pumps of different efficiency are available, e_1 is a convex function.^{2.2}

Example 2: Thermal Storage

In many situations the temperature of some area must be kept within acceptable limits, using the thermal mass of the area for storage. Simple versions of this can be modeled as a one-stage process, with $S_{\max,1}$ the difference between maximum and minimum allowable temperatures times the thermal mass, $X_{\max,1}$ the maximum heating or cooling rate, $C_1 = 0$, e_1 the efficiency (in joules per kWh), L_1 the losses due to heat transfer to the environment; and

$X_{\min,2} = X_{\max,2}$ the loss when storage is empty, which will usually be non-zero.^{2,3} The system is operated by turning the heating or cooling units on full for a few minutes, then turning them off. Again, $S_{\max,1}/X_{\max,2}$ is a critical parameter. Here it measures how long electricity can be turned off without violating allowable temperature constraints.

Examples of thermal storage include space heating, air conditioning, refrigeration, food freezing, and electric furnaces. For those carried out near ambient temperature, weather will make a large difference in how much demand can be rescheduled. For example, for air conditioning a commercial building, hot weather will increase $X_{\min,2}$. This raises total demand over a day; it also reduces the time that chillers can be turned off.

Most thermal systems have a few minutes to a few hours of storage under normal circumstances. Under time-varying prices (either spot or predetermined), the economic value of storage increases tremendously. This suggests new optimal designs for the system, such as installing chilled water tanks to increase $S_{\max,1}$.

Example 3: Steel Mini-Mill

Small steel mills based on electric arc furnaces use several megawatts of electricity each. Steel is processed through the arc furnace in batches, each taking several hours [Gordian Associates, 1980, p. III-4]. A typical steel mill has other production stages which are less electricity-intensive. Their schedule can be treated as fixed. Such mills can be modeled as storage/flow processes at several levels.

First, output per day and all labor schedules can be held constant but production moved around within the day. Optimal behavior would be to shut down the furnace during peak electricity prices.^{2.4}

Although storage capacity will not be a constraint, furnace capacity (melts per day) may be if the plant is operating at full capacity. More flexibility can be achieved by rescheduling production between days of the week.

Second, labor scheduling could be adjusted to have more production at night. The attractiveness of this depends on the kilowatts per worker shifted. Once adopted, this cannot be easily changed from one day to the next. Therefore evaluating the economics of this alternative will require forecasts of $p(\text{daytime}) - p(\text{nighttime})$ many months ahead. In effect, any labor change is a capital investment. Electricity use can be rescheduled by roughly one-half day using this method.

Third, scheduling of labor and production over the day could be held fixed, but the heating elements cycled differently from minute to minute. Cold raw materials are added and heated until they melt. Once they melt, the temperature can be allowed to cycle within a band during the one- to two-hour alloying process. This is an example of thermal storage. Maximum rescheduling is roughly half an hour.

All of the above are examples of pure storage behavior. That is, the firm's final production level is unchanged over the course of the planning period. Electrical use is moved around but total consumption is approximately constant. At the opposite extreme is pure shutdown

behavior. For the steel mill, this would be simply to not make steel when the electricity price is too high, and never make up the shortfall. As I will show in the next section, optimal storage behavior is determined by relative prices at different times, while shutdown behavior depends on absolute prices.

Example 4: Fossil electric generation is a one-stage "pure shutdown" process, with $S_{\max,1} = 0$, $C_1(t)$ the cost of fuel per Btu, $X_1(t)$ the rate of fuel use, and negative electricity "use" $e(X_1)$. Thus the firm will turn on or off whenever the electricity price crosses the marginal generating cost.^{2.5} I will study a diesel electric generator in Chapter 5.

Example 5: Run-of-river hydro generators have zero variable costs $C_1(t)$. $X_{\max,1}(t)$ represents the inflow of water and $X_2(t)$ the outflow. Many sites have ponds which can be used to store a small volume of water, corresponding to $S_{\max,1}$. Given time-varying prices the owner would have an incentive to adjust $X_2(t)$ appropriately. I will study this in Chapter 5.

Conclusion

A great variety of uses and generating methods for electricity can be modeled as storage/flow processes. Electricity can be "stored," embodied in intermediate products. Such storage has costs and constraints which are process-specific, and is therefore mainly

feasible under decentralized control by end users. Such "storage" is also usually irreversible, but this does not reduce its value to appropriate end users.

"Storage" and "shutdown" are two different modes of response to spot prices with rather different properties. Centralized and decentralized generating technologies can be modeled as either pure storage or pure shutdown processes, depending on the technology.

4.3 Properties of Optimal Behavior

The system (4.1.1) and (4.1.2) can be used to predict the behavior of a specific firm under a specific set of electricity and other prices. This requires measuring the levels of each kind of capital stock, which determine the production and storage capacity constraints (4.1.2b and 4.1.2e), the shape of the electrical use functions $e_n(X_n)$, and other relevant information. This approach will be followed in Chapter 5.

This section instead looks at the properties of solutions as a function of electricity prices. It shows:

- o How optimal electricity demands in each period depend on prices in that and other periods, and on the amount of foresight about future prices.
- o How optimal electricity demands depend on the capital stocks $S_{\max,n}$ and $X_{\max,n}$.
- o How the incremental values of additional investments depend on the rate structure.
- o How the value of the objective function depends on the rate structure. The objective function is $F_j(D_j, \underline{K})$ in Chapter 3, where $D_j(t) = \sum_n X_n(t)$ for firm j.

Thus the model of this chapter provides a tool to evaluate the derivative of F_j with respect to various quantities. These derivatives are needed for optimal investment planning and other purposes.

For easier analysis I concentrate on a two-stage customer. As shown in Section 4.2, this is adequate to capture important behavior for many firms. Multi-stage models show similar behavior, but it is "diluted" by interactions among stages.^{3.1}

Deterministic Behavior

I will start with the deterministic case, where $p(t)$ is either a time-of-use rate, or a spot rate with prices revealed for certain by time 0. For the resulting problem we get the Lagrangian:

$$\begin{aligned}
 (4.3.1) \quad & \text{Cost of} & \text{Cost of} & \text{Value of} \\
 & \text{other} & \text{electricity} & \text{product } (<0) \\
 & \text{inputs} & & \\
 \text{Min } & \sum_{t=1}^T C_1(t)X_1(t) + p(t)e_1(X_1(t)) + C_2(t)X'_2(t) \\
 & + \alpha_1(t)[X_1(t) - X_{\max,1}] \\
 & + \alpha_2(t)[X'_2(t) - X_{\max,2} + X_{\min,2}(t)] \\
 & + \beta_1(t)[S_1(t) - S_{\max,1}] \\
 & + \gamma_1(t)[S_1(t) - L_1S_1(t-1) - X_1(t) + X'_2(t) + X_{\min,2}(t)]
 \end{aligned}$$

where $X'_2(t) = X_2(t) - X_{\min,2}(t)$ = discretionary final production

$$\text{s. t. } 0 \leq X_1(t) \leq X_{\max,1}$$

$$0 \leq X'_2(t) \leq X_{\max,2}$$

$$0 \leq S_1(t) \leq S_{\max,1}$$

$$S_1(0) \text{ given.}$$

Here α , β , and γ are the appropriate dual variables.

The Kuhn-Tucker and complementary slackness conditions for this problem are given in Figure 4.3.1. From these conditions we can derive optimal behavior and the shadow prices on capacity. However a graphical approach is much easier.

I will make several simplifying assumptions and definitions which lead to a problem that can be solved graphically. An appendix proves the correctness of the graphical solution, using a continuous time model.

- o The value of final production is constant, $C_2(t) = C_2$.

It must be that C_2 is negative so that the value of the final product is positive.

- o Minimum production $X_{\min,2}(t) = 0$.^{3.2}
- o Marginal electricity consumption is a constant

$$\frac{\partial e_1(X_1)}{\partial X_1} = e_1' = \text{constant} = \text{electricity use in kwh per unit of production.}$$

- o Storage losses are zero: $L = 1$.^{3.3}
- o $S_1(0) = 0$. No initial storage.

We can now find the total variable cost of producing a unit of X_1 at time t . It is $X_1(t) [C_1(t) + e_1' p(t)]$. Define

$$(4.3.5) \quad P_1(t) = \left[\frac{C_1(t)}{e_1'} \right] + p(t)$$

Then the process can be modeled as if electricity is the only variable input, and has a "price" of $P_1(t)$. Notice that the importance of electricity price in determining optimal behavior depends on

Figure 4.3.1
Conditions at Optimum

(4.3.2) Upstream stage output	<u>Relation</u>	<u>If</u>
$0 < C_1(t) + p(t) \frac{\partial e_1}{\partial X_1(t)}$	= $-\alpha_1(t) + \gamma_1(t)$	$X_1(t) = X_{\max,1}$
	= $\gamma_1(t)$	$0 < X_1(t) < X_{\max,1}$
	> $\gamma_1(t)$	$0 = X_1(t)$
(4.3.3) Final production	<u>Relation</u>	<u>If</u>
$0 \leq -C_2(t)$	= $-\alpha_2(t) + \gamma_1(t)$	$X'_2(t) = X_{\max,2}$
	= $\gamma_1(t)$	$0 < X'_2(t) < X_{\max,2}$
	< $\gamma_1(t)$	$0 = X'_2(t)$ (shutdown)
(4.3.4) Storage level*	<u>Relation</u>	<u>If</u>
$0 = -\gamma_1(t) + L_1\gamma_1(t+1) - \beta_1(t)$	$S_1(t) = S_{\max,1}$	(Storage full)
$0 = -\gamma_1(t) + L_1\gamma_1(t+1)$	$0 < S_1(t) < S_{\max,1}$	(Storage charging or discharging)
$0 = -\gamma_1(t) + L_1\gamma_1(t+1)$	$0 = S_1(t)$	(Storage empty)

*If there is an inventory holding cost per period, its derivative per unit in storage goes on the left-hand side of equation 4.3.4.

$C_1(t)/e_1'$. If variation in $C_1(t)$ over time is large relative to variation in electricity prices $p(t)$ over time, or if e_1' is small, then the firm will not be very responsive to electricity prices.

Optimal behavior is clearest for the simplest interesting price pattern: a low price which rises and then falls again. Any vector of prices over a day consists of one or more such patterns.^{3.4} For now assume that a day has only one price peak, as in Figure 4.3.2.^{3.5}

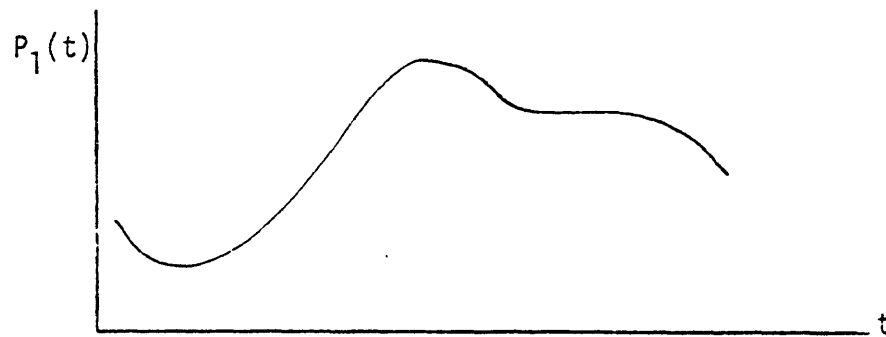


Figure 4.3.2

"One Price Peak" Case

Optimal behavior for producing $X_1(t)$ over this price cycle is, to the extent allowed by constraints, to:

- o Charge up storage when the cost is low, by running at $X_1(t) = X_{\max,1}$.^{3.6}
- o Shut down stage 1 ($X_1(t) = 0$) when the cost is near its peak.
- o Start stage 1 back up ($X_1(t) = X_2(t)$) when storage is exhausted.
- o Set discretionary final production $X_2(t)$ to zero whenever storage is empty and $P_1(t)e_1'$ is higher than the value of final product $|C_2|$.
- o At other times set discretionary final production to its

upper limit $X_{\max,2}$.

The extent to which this behavior can be followed depends on how much slack exists in stage 1 production capacity, and on how much storage exists. Two physical parameters provide sufficient statistics to describe this production process and measure its slack: the number of hours to empty and to fill storage when running at full capacity.

Define

$$(4.3.6) \quad s_1 = S_{\max,1}/X_{\max,2} \quad \text{hours to empty storage}$$

$$x_1 = S_{\max,1}/(X_{\max,1}-X_{\max,2}) \quad \text{hours to fill storage}$$

s_1 is the number of hours to empty storage when $X_1 = 0$ and X_2 is at maximum; x_1 is the number of hour to fill storage when X_1 and X_2 are both at their respective maxima.^{3.7}

Given prices over time, storage parameters s_1 and x_1 , and the value of final production $|C_2|$, optimal behavior can be determined. Thus a small firm with four hours of storage and 2-hour charging time will act like a scaled version of a much larger electricity user with the same parameters.

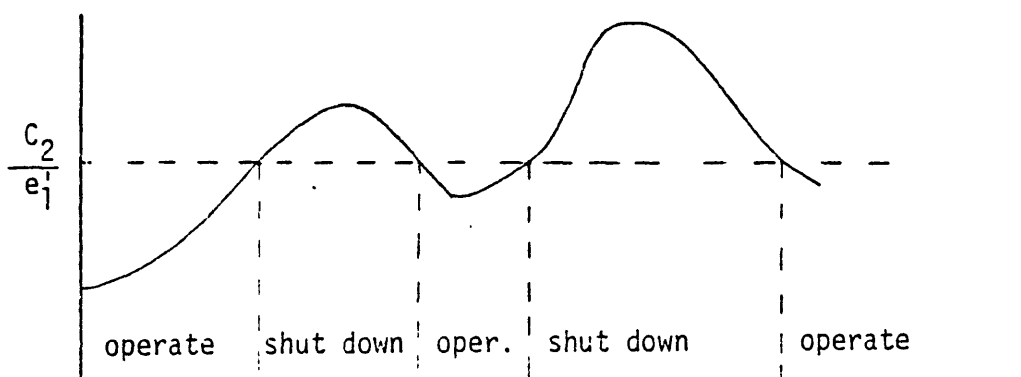


Figure 4.3. 3

Pure Shutdown Case: Graphical Solution

Figure 4.3.3 shows the "pure shutdown" case of a process with no storage capacity, i.e., $s_1 = 0$. Optimal behavior is to shut down completely whenever the "electricity" price $P_1(t)$ is above the value of the final product $|C_2|$. Thus demand is infinitely price-responsive at $|C_2|$.^{3.8} The effect on the customer is to "chop the peaks" off the prices $P_1(t)$. At each period the customer has the option to shut down, at an exercise price of $|C_2|$. If there is any probability it will want to exercise, then the option is valuable.^{3.9} The value of the option is the area under the price duration curve above the shutdown point. This was discussed in Section 3.1 for generators.

If any storage of final or intermediate product is possible, a one-period demand curve is inadequate. Optimal behavior at each time depends on prices at other times. Figure 4.3.4 shows a "pure storage" case, in which $P_1(t)$ is below $|C_2|$ at all times. The firm will "charge up" storage between 0 and T_f , and discharge from T_d to $(T_d + s_1)$, thus following the rules above.

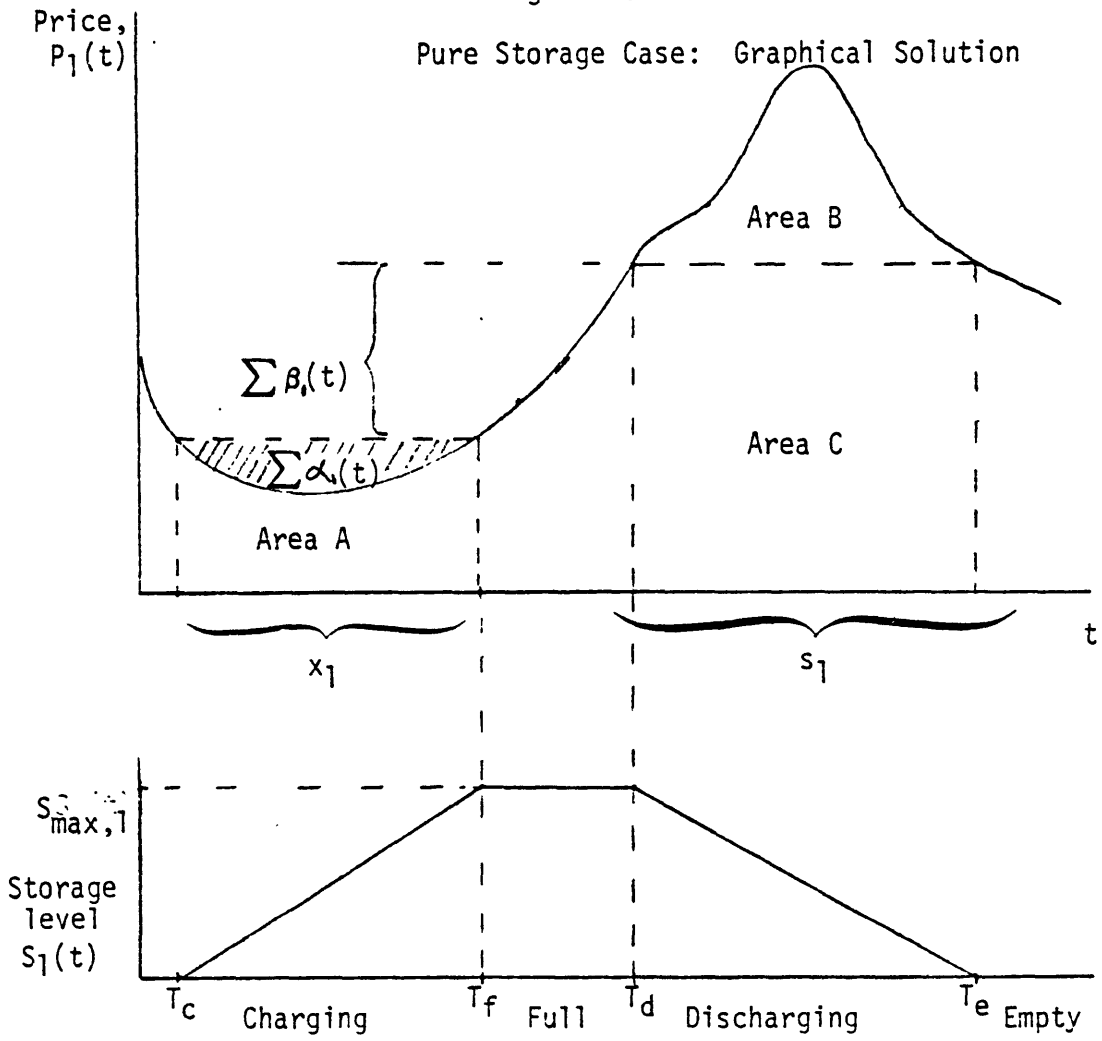
Value of Investments

$\beta_1(t)$ measures the shadow value of storage capacity in period t . The value of 1 kwh of additional storage capacity over a cycle is

$$\sum_{t=1}^T \beta_1(t)$$

This is a very important value for marginal investment decisions.

Figure 4.3.4
Pure Storage Case: Graphical Solution



Marginal value of stage 1 production capacity: $\sum \alpha_i(t)$
 Marginal value of storage buffer between stages: $\sum \beta_i(t)$

Total value of both = $[\text{Area B} + \text{Area C}] - (\text{Area A}) \times \frac{\lambda_{\max,1}}{\lambda_{\max,2}}$

Comparing its expected value with the amortized capital cost of further investment in storage determines

whether such investments should be made. Similarly, $\sum_{t=1}^T \alpha_1(t)$ is the total value over a cycle of investments in increased stage 1 capacity.^{3.10} Figure 4.3.4 shows the values of $\sum \beta_1(t)$ and

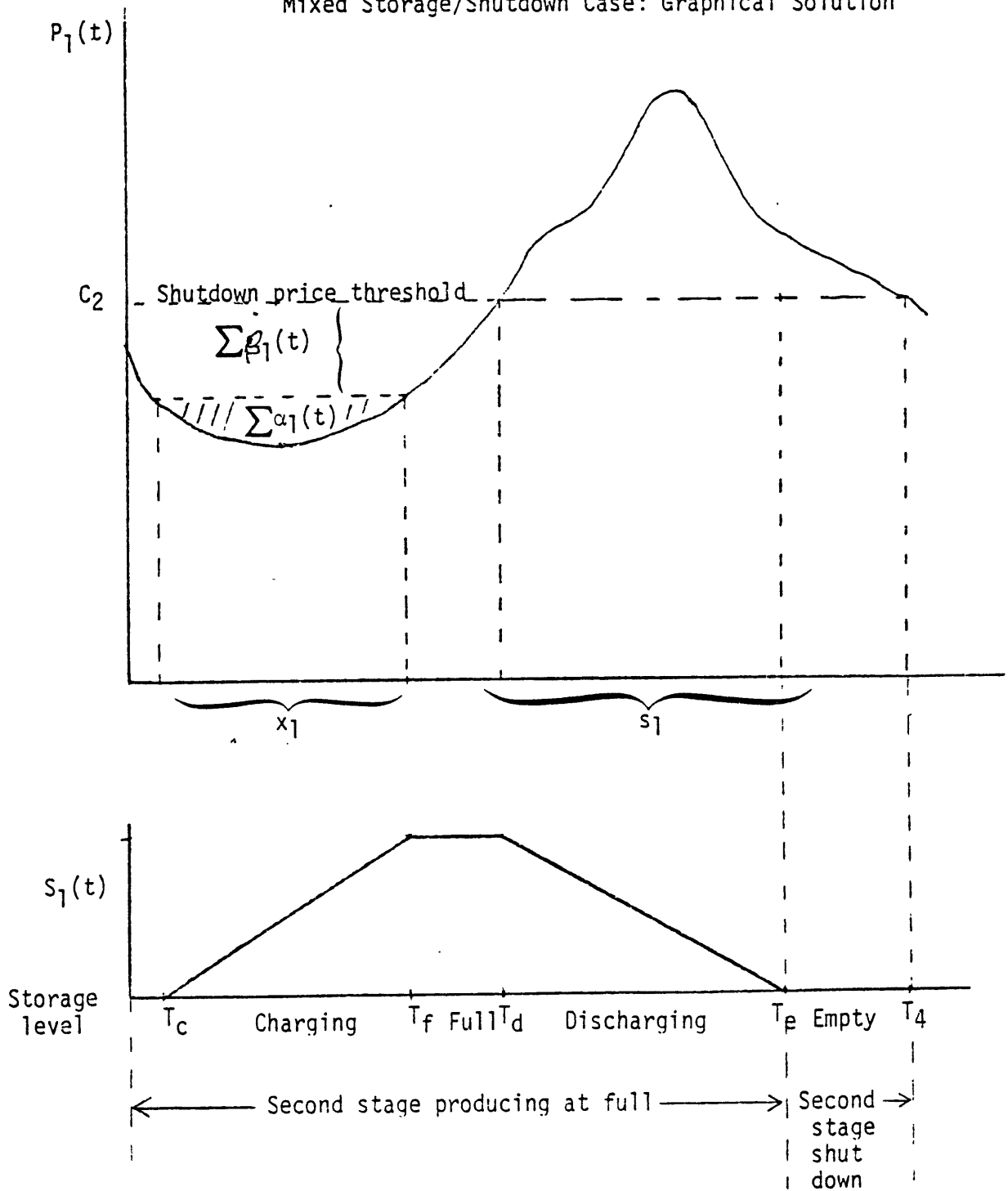
$\sum \alpha_1(t)$, assuming optimal behavior. Notice that the marginal values of storage aiding investments are increasing functions of the magnitude of price changes over the relevant periods.^{3.11} Only price differences, not absolute prices, matter in the pure shutdown case.

The level of existing capital stock also affects the value of capital investments. As charging time x_1 grows, i.e., production capacity $X_{\max,1}$ falls, the value of a unit of increased stage 1 capacity (α_1) rises; but the value of increased stage 1 storage (β_1) falls. Similarly, a decrease in $S_{\max,1}$ (which lowers storage time s_1 and x_1) will decrease the value of incremental stage 1 capacity but raise the value of increased storage capacity. Many production processes today are built with little "excess" capacity, hence have high shadow prices for such capacity under time varying prices.^{3.12}

Figure 4.3.5 shows the same production process under higher prices. The prices rise enough to make some shutdown of stage 2 optimal, as shown at the bottom. By T_e all storage has been exhausted and $e^{-1}P_1(t) > |C_2|$, making production uneconomic. The shadow value of incremental storage capacity β_1 falls to almost

Figure 4.3.5

Mixed Storage/Shutdown Case: Graphical Solution



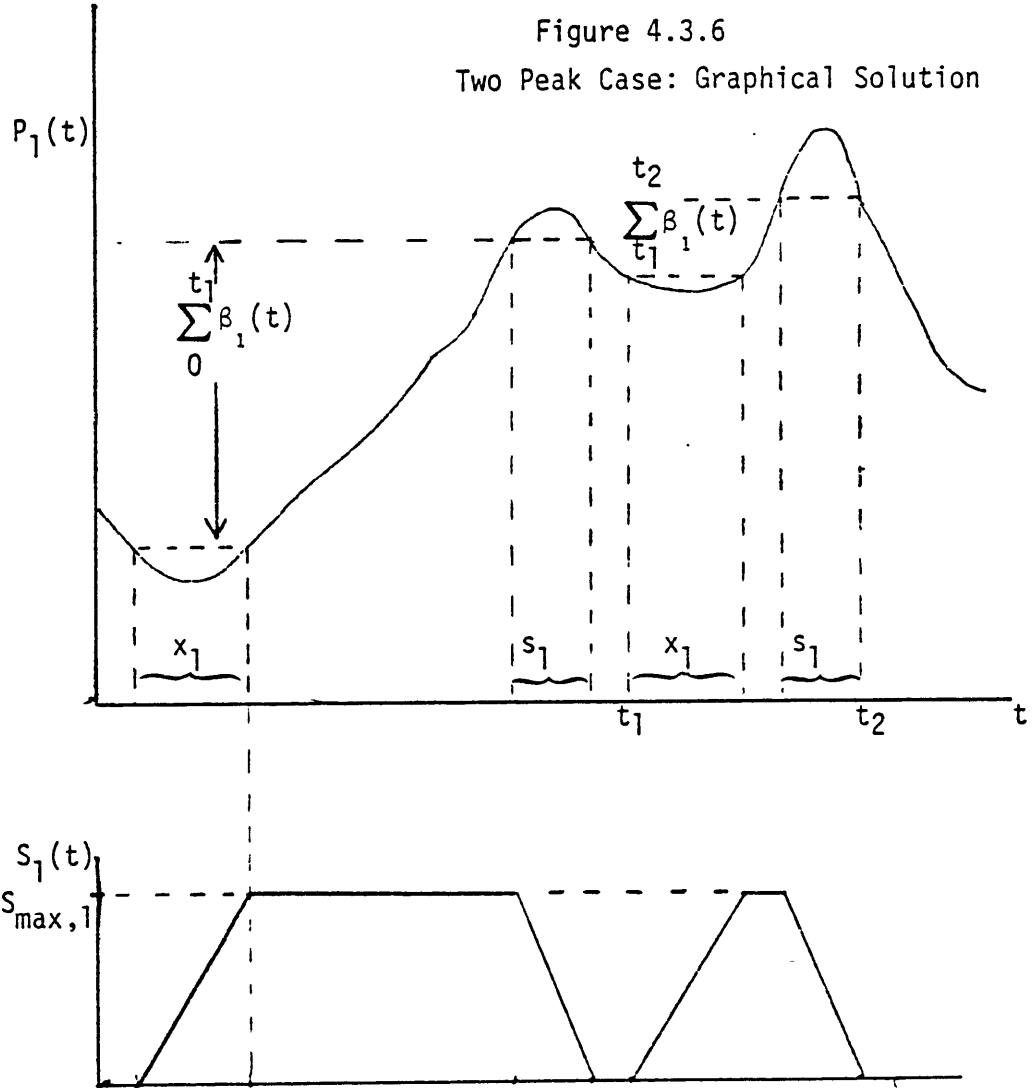
zero, because discharge begins at almost the same price at which charging ended.^{3.13,3.14}

Figure 4.3.6 shows what can happen if prices for the day have two peaks. Behavior during each part of the day can be analyzed as if it were a single peak cycle. Then there is a consistency relation between the two days which must be met.^{3.15} In the example shown, s_1 is small relative to the width of the peaks, and x_1 is small relative to the valley between them. Therefore the peaks can be analyzed separately. Hence the total value of storage over the cycle is the sum of its values for each trough to peak episode.^{3.16}

Stochastic Case

Behavior when future prices are stochastic must be found by taking the current state of the system as given and minimizing expected future costs, given current knowledge about future prices' probability distribution. After period t the state of the system is $S_1(t)$. The value of the costs from $t + 1$ to T is a decreasing function of this state. Optimal behavior in t is to minimize the expected value of these costs plus the cost of going from $S_1(t - 1)$ to $S_1(t)$. This dynamic programming problem will be formulated as a recursive expansion of the deterministic L.P. I will then discuss behavior under stochastic prices. The firm can always follow adaptive certainty equivalent behavior; some firms can do better. Advance warning about future prices helps the firm respond better. Nonetheless, predetermined prices are not better for the firm than

Figure 4.3.6
Two Peak Case: Graphical Solution



spot prices with the same expected values.

Reformulate the Lagrangian (4.3.1) for the stochastic case by defining the information set $I(t)$ as all new information arriving at the beginning of period t . One element of $I(t)$ is the exact value of $p(t)$. Another may be information about the probability distribution of future electricity prices. Optimal control actions are functions of the system state and of the information set:

$$X_1(t) = X_1(t, I(t), S_1(t))$$

$$X_2(t) = X_2(t, I(t), S_1(t))$$

Now consider the firm's problem at time s . It wishes to minimize the conditional expectation of costs from s to T , knowing that actions at $s + 1, \dots, T$ will depend on current action and on to-be-revealed information $I(s + 1), \dots, I(t)$.

Assume that the firm has some joint probability distribution for future information, conditional on past information. Its problem is to select $X_1(s), X_2'(s)$ to minimize:^{3.17}

$$(4.3.7) \quad P_1(s)X_1(s, I(s), S_1(s - 1)) + C_2X_2'(s, I(s), S_1(s - 1))$$

$$+ \alpha_1(s)[X_1(s) - X_{\max,1}]$$

$$+ \alpha_2(s)[X_2'(s) - X_{\max,2} + X_{\min,2}(s)]$$

$$+ \beta_1(s)[S_1(s) - S_{\max,1}]$$

$$+ \gamma_1(s)[S_1(s) - S_1(s - 1) - X_1(s) + X_2'(s) + X_{\min,2}(s)]$$

$$\begin{aligned}
& + E_s \sum_{t=s+1}^T P_1(t) X_1(t, I(t), S_1(t-1)) + C_2 X'_2(t, I(t), S_1(t-1)) \\
& + \alpha_1(t) [X_1(t) - X_{\max,1}] \\
& + \alpha_2(t) [X'_2(t) - X_{\max,2} + X_{\min,2}(t)] \\
& + \beta_1(t) [S_1(t) - S_{\max,1}] \\
& + \gamma_1(t) [S_1(t) - S_1(t-1) - X_1(t) + X'_2(t) + X_{\min,2}(t)]
\end{aligned}$$

s. t. non-negativity constraints and upper bounds;

initial storage $S_1(s-1)$ given.

where $E_s =$ expectation given all past information $I(s)$,
 $I(s-1), \dots, I(0)$.

All primal and dual variables are functions of realized past information and anticipated future information.

Optimal future actions $X_n(t)$ will depend on $S(t-1)$ which in turn may depend on future information $I(s), I(s+1), \dots, I(t-1)$. If $I(s)$ reveals prices until the end of the cycle with certainty, (4.3.7) is just a subset of the deterministic problem (4.3.1), solvable by graphical or L.P. methods. Otherwise, optimal $X_n(s)$ will depend on the probability distribution of future prices.

At worst, the firm can always solve the certainty equivalent problem formed by estimating expected prices

$$(4.3.8) \quad p'(t) = E_s p(t)$$

and assuming $p'(t)$ for certain. If the firm chooses at s to "lock in"

future $X_n(t)$, the expected value of its objective function is (by linearity) the same as if the prices $p'(t)$ were charged for certain. However the firm may do better by following adaptive behavior: choosing only $X_1(s)$, $X_2(s)$ now and assuming (4.3.8), then repeating the procedure next period using estimates of $p'(t)$ updated by the arrival of $I(s + 1)$.

Thus a risk neutral firm will always have expected profits which are as high under spot pricing as under predetermined prices set to the ex ante expected values. It can do better if there is some positive probability that new information will arrive which makes it optimal for the firm to change its initially chosen behavior. A firm with a flexible production process owns various options to change its behavior. Spot prices are valuable if they have a wide enough range that the firm might choose to exercise these options. The more likely to be exercised, the more valuable the option. Notice that the result applies to any two rates, one of which is recalculated more often than the other. For example, rates with an annual recalculation are no better and probably worse for the risk neutral firm than rates with daily recalculation. Daily recalculation is in turn dominated by hourly recalculation, for a firm flexible enough to profitably adjust its production plans with less than a day of advance warning.

Although this result is interesting and, to many, counter-intuitive, it does depend on the specific assumptions used to prove it. The result arises from the assumption that the firm has a technology describable by equations 4.1.1 and 4.1.2, which give it

convex profits as a function of spot prices $p(t)$. It is closely related to the social value of putting a firm on another rate, discussed in Section 3.5. As we saw, if the firm receives a subsidy under \underline{p} it may not prefer spot prices. The price assumption (4.3.8) here rules out subsidies due to time aggregation of rates. And the production technology of this chapter assumes away subsidies due to aggregation over states of nature, since demands depend only on price. Hence both terms of the subsidy equation (3.5.4) are zero. More generally, of course, they may be non-zero, with the implication that self-selection between the rates may be socially suboptimal.^{3.18}

Although adaptive certainty equivalent behavior under spot pricing will give most firms better results than will predetermined prices, certainty equivalent behavior is not necessarily optimal under spot prices. That is, the optimal solution of (4.3.7) may lead to behavior at time s which depends on more than $E_s p(t)$. In general, optimal behavior will depend on the rank ordering of prices in different periods.

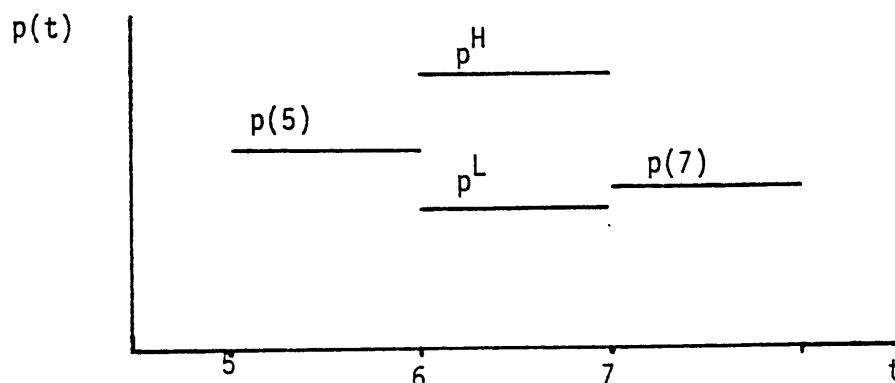


Figure 4.3.7

Non-Optimality of Certainty Equivalent Behavior

Figure 4.3.7 shows an example. Suppose $S_1(4) < S_{\max,1}$ and $p(6) = p^H$ with probability q or p^L with probability $(1 - q)$. Should storage be charged during period 5? The answer depends only on $p(5) - [q p^H + (1 - q)\text{Max}(p^L; p(7))]$

If this is positive, it is optimal not to charge.^{3.19} If negative, storage should be charged as much as possible. But if $p^L < p(7)$ the critical expression does not depend on p^L , hence cannot depend on $E p(6)$.

This failure of certainty equivalence arises because the control variables $X_n(t)$ have a nonlinear impact on the state variables. Once a bound is reached, increasing a control has no effect. Hence the stochastic case cannot be correctly modeled as a series of deterministic models, contrary to the claims by Takayama and Judge [1971, Ch. 19] and others.

Other Effects of Uncertainty

Several other important results can be derived from the nested models approach of equation (4.3.7).

- o The amount of advance warning about future prices affects both behavior and expected profits. The earlier future prices are known, the better off the firm is, and the more responsive $X_n(t)$ will be to actual prices $p(t)$. Reducing uncertainty about prices to zero is not necessary; optimal behavior can be determined if the ordering of future prices is known with certainty, without knowing the exact

levels.^{3.20} And once ex post optimal behavior is known with certainty, further information is irrelevant.^{3.21}

Thus the best utilities for spot pricing are those with a large amount of initial variability about full spot prices, but an early resolution of all uncertainty about future spot prices.

That is, a high unconditional forecast variance of future spot prices is desirable since it increases the value of the firm's various options to adjust behavior to prices, by shutting down, storing, or whatever. So is a high inter-period variation. But a low forecast variance conditional on information available a month in advance is also desirable, since it gives the firm more time to make adjustments, and increases the probability they will be ex post optimal. Given a choice among various price patterns with the same mean price, customers would prefer a completely predictable square wave with high amplitude and short period. Storage customers would save by storing each time prices were low. Shutdown customers would save each time the price exceeded their shutdown point. The best practical case is probably a utility with large but predictable demand fluctuations.

Variability and uncertainty in spot prices affect the values of production and storage capacity in ways similar to the above effects on expected profits. This is important because it determines firms' incentives to invest.

- o More deterministic variability in spot prices over time increases the value of investments. For example, the shadow value of storage capacity $\sum \beta_n(t)$ is given by the difference between the prices at which charging stops, and discharging begins. (See equation 4.3.4 or Figure 4.3.5. A proof is in the appendix.) Thus the larger the amplitude of price fluctuations, the more valuable is additional storage.
- o Decreases in the period of price cycles, however, decrease the marginal values of production and storage capacity. The reason is obvious: Going from n to $n+1$ hours of storage capacity has no value when a price peak lasts fewer than n hours. Thus incremental storage has a higher value to a firm under two-level time-of-use prices, than under hourly spot pricing, even though the total value of its storage is higher under hourly spot pricing. (See Chapter 5 for examples.)
- o Stochastic variation in spot prices is not quite as clear-cut. I conjecture but have not proved that it too will increase the value of incremental storage. Specifically, given the pricing systems $p^*(t)$ and $p'(t)$, if they have the same unconditional expected values at each moment but the probability distribution of $p^*(t)$ is wider, then the expected incremental shadow prices $E \sum \beta_n(t)$ and $E \sum \alpha_n(t)$ are at least as high under pricing system $p^*(t)$ as under $p'(t)$. All cases I have looked at have this property, and one can construct an intuitive argument for it.^{3.22}

- o If two pricing systems have different means at different times, then it is possible that the one with less uncertainty will still give a higher expected incremental value to investments. Thus certain investments may be more profitable under rates with time aggregation error than under full spot prices. An example of this occurs in one of Chapter 5's case studies. However, the investment had lower social value under time aggregated rates.
- o As is true for expected profits, better advance knowledge about future spot prices will increase the expected incremental values of capacity, since it allows the capacity to be used more efficiently. Again this does not imply that artificially stabilized prices will lead to higher values, only that accurate price forecasts lead to higher values of capacity than inaccurate forecasts.

Conclusion

Spot prices lead to higher expected profits and higher customer incentives to invest than other rates, as long as the other rates do not have built in cross-subsidies for the customer.^{3.23,3.24}

Also full spot prices ensure that the profit-maximizing behavior discussed in this section is also social welfare-maximizing.

Adaptive certainty equivalent behavior, i.e., acting as if future prices are known for certain but revising this "known" future each time new information arrives, is one behavioral mode for customers on spot pricing. However, it is possible that the customer can do better by following a more sophisticated strategy.

4.4 Experience with Spot Prices

Full spot pricing has not been implemented by any public utility that I know of. But many utilities have limited versions of stochastically changing prices. A common version of stochastically time-varying prices is the monthly or quarterly fuel adjustment charge for electricity and gas sales. As shown in Figure 2.0.1 these can lead to 20 percent price changes from one period to the next. As far as I know, no one has studied demand response to these prices.^{4.1}

The major systematic uses of stochastic prices are in Great Britain, documented in Acton et al. [1980] and Mitchell et al. [1979] and in Sweden [Camm, 1981]. San Diego has a mandatory "coincident demand charge" for large customers. Illinois Power is implementing a voluntary "interruptible rate" which is really a stochastic price, since customers can avoid interruption for a fee per kilowatt hour used [see Gorzelnik, 1980].

Experience with the San Diego rate qualitatively confirms the models of this chapter. Customers exhibit both storage and shutdown behavior in various ways. Some customers, however, apparently do not respond to spot prices at all. This indicates either a very high shutdown price threshold for their operations, or that the chapter overlooks important issues. Bohn [1980] and Bohn [1981] discuss in more detail my findings from interviews and econometric analysis.

Firms in Great Britain are quite responsive to high spot prices, which they are warned about several hours in advance.^{4.2}

Notification is by telex, and designates several hours which will be a

peak period. Customers can still use electricity, but pay a high demand charge for their use during this period.^{4.3}

Electricity-intensive firms respond strongly to the high prices. Examples discussed by Acton et al. fit the models of this chapter.

They include:

- o Reducing total production by chlorine producers, i.e., shutdown behavior [Acton et al., p. 24].
- o Increased generation [p. 24, 26, 42, 45] by diesels and steam based cogeneration. In at least one case, a cogenerator wasted steam to produce maximum electricity.
- o Storage and shutdown behavior for the final products by gas liquefaction plants.
- o Sequential shutting down of a cement plant, except for its kilns [p. 34].
- o Scheduling meals and breaks during high price periods, and making up lost production later [Steel mill, p. 51].

Sweden has a system which is very close to full spot pricing for about 15 very large industrial users and self-generators [Camm, 1981]. Apparently they do not use the optimal spatial price differences derived in Chapter 3, but prices are instead basically equal to system λ . There is tremendous variation in spot prices over a month [p. 42]. Industrial firms with cogeneration respond to price variations when the prices are extreme, in some cases shutting down their cogenerators and using electric boilers for steam when prices are low enough [p. 68].

4.5 Conclusion

The thrust of this chapter is that behavior of most large electricity users or generators under time-of-use or spot prices can be modeled before such rates are implemented. From this analysis we can hypothesize the effects of spot pricing on different firms. How many and what kinds of actual firms will be most responsive to and derive the most benefit from spot prices? What spot prices will give the largest responses? How will firm response affect the level of generation, costs and spot prices, if full spot pricing is implemented?

Suitable Industries

Figure 4.5.1 summarizes characteristics of firms and processes within firms which could be most responsive to spot prices. Based on Figure 4.5.1 and the earlier discussion we can look at different industries to see which are most likely to be suitable. These suitable industries include:

- o Those involving pumping a liquid into storage tanks, such as municipal water systems, oil pipelines, and coal slurry pipelines.
- o Processes where liquids are pumped directly to a processing area, but residence time in the processing area is not critical. Agricultural irrigation is one such case.
- o Industrial gas liquefaction plants.

- o Air conditioning and refrigeration processes, especially where a storable heat transfer fluid is used.
- o Heating processes where the end use costs of fossil fuel and electricity are close to each other. Food and crop drying are examples. This process is used for one of Chapter 5's case studies.
- o Firms with multiple alternate ways of cogenerating steam and electricity. Cogenerators with backup low-pressure boilers are an example. Some pulp and paper mills fit this description.

Many of these firms and industries are discussed in Acton et al. [1980] or Gordian Associates [1980]. I return to them in Section 5.6, where I make a preliminary assessment of their size.

Figure 4.5.2 summarizes the characteristics of prices which will tend to increase the profits of responsive firms. In Section 5.5 I discuss what kinds of utilities will have full spot prices with these characteristics.

Market "Equilibrium" with Full Spot Pricing

User demands and generation capital stock and operating behavior combine to determine optimal full spot prices, as detailed in Chapter 3. Actually applying full spot prices to participants will alter

Figure 4.5.1
 Characteristics of Processes Which Will Be Most Responsive
 to Spot Prices*

- o Total demand for final output often less than maximum production capacity. Thus final goods inventory can be used for storage, and upstream shutdowns can be long enough to reduce final output. This pattern is characteristic of some processes and industries with fluctuating demands, such as primary metals.
- o Easily storable intermediate products, allowing low-cost expansion of storage buffers. (For example hot water, but not steam).
- o Electricity costs a large fraction of value added. This makes reducing final output, i.e., shutdown behavior, more attractive.
- o Several production stages with one stage quite electricity-intensive.
- o Large intermediate-product storage buffers before and after the critical stage. Also low leakage (L_n close to 1.0) and low carrying costs for inventory.
- o A critical stage which usually has excess capacity.^{5.1}
- o Low labor intensity in the critical stage. Or, a highly flexible labor force which can be shifted to other tasks with almost as much value added per man hour. (Example: a flexible maintenance schedule.)
- o Automated or simple process control, which reduces the costs of rescheduling production from one cycle to the next (Section 5.4).

*Most of these characteristics also predict which firms will be responsive to time-of-use prices.

Figure 4.5.2
Characteristics of Prices Which Increase Firms' Profits,
Price Responsiveness, and Value of Additional
Storage/Production Capacity

- o Large and frequent changes in price over time.
- o Large stochastic variability of spot price at a particular hour. (Variable weather and other demand influences; large power plants.)
- o Early resolution of actual spot prices. (Predictable weather and demand; reliable scheduling of generator maintenance.)

their behavior and therefore alter the prices. The effect will be to dampen but not eliminate both predictable and unpredictable price fluctuations over time. Adjustment to spot prices will be gradual, as customers and independent generators make appropriate investments.

In classic models of commodity pricing over time, expected prices must follow a steady price path, such as the classic Hotelling rule that price increases at the rate of interest. The shape of this path is essentially independent of the predictable component of demand curve fluctuations. The reason is that producers will stockpile production if price is expected to increase faster than that path, thus forcing up prices now. Conversely, speculators will sell short if they anticipate slower increases. The equilibrium expected price path is determined by inventory holding costs.^{5.2} In these models new information about present or future demands or supplies leads to discontinuous adjustments in price. Price changes then return to a monotone path, starting from the new level. Wide anticipated swings in prices are not possible.

Full spot prices for electricity under current conditions do show regular diurnal cycles in expected prices. (See Chapter 5 for examples.) The reason is that storing electricity is expensive. Therefore utilities presently build only limited amounts of centralized storage. That storage dampens predictable price cycles, but only a little.

Implementing full spot pricing to customers will diminish the amplitude of these cycles but not eliminate them. Spot pricing will

give customers economically correct incentives to make use of existing storage capacity and to build more storage. Since decentralized storage uses totally different technology and storage media than conventional storage, many potential investments exist which are cheaper than conventional storage and would be economic under full spot pricing. Building and operating decentralized storage in response to spot prices will therefore decrease the magnitude of full spot price fluctuations, as shown in Figure 4.5.3. This decreases the value of further investments in storage. Such investments have increasing marginal costs across and within firms, since some processes are more suited to storage than others. Therefore equilibrium is reached with some, and perhaps still large, daily anticipated cycles. 5.3,5.4

Spot pricing will also dampen price fluctuations by encouraging "shutdown" behavior in the short run and investments in shutdown equipment (such as standby generators and easier-to-shut-down production technologies) in the long run. Since some shutdown methods need essentially no advance warning, they will be effective at damping unanticipated peaks in spot prices. Storage techniques will be effective mainly at damping price rises which are anticipated in advance. 5.5

The impact of spot pricing will vary depending on business conditions. In classic commodity models, storage is reversible and not associated with a particular process. Therefore the costs to speculators of storing to meet an anticipated price rise are not

Figure 4.5.3

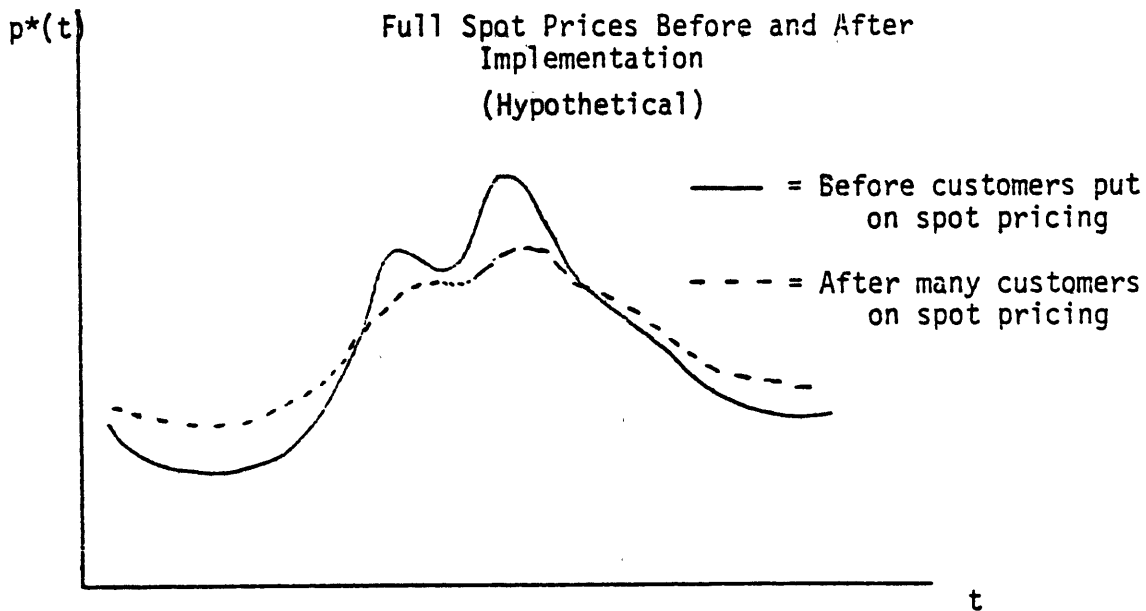
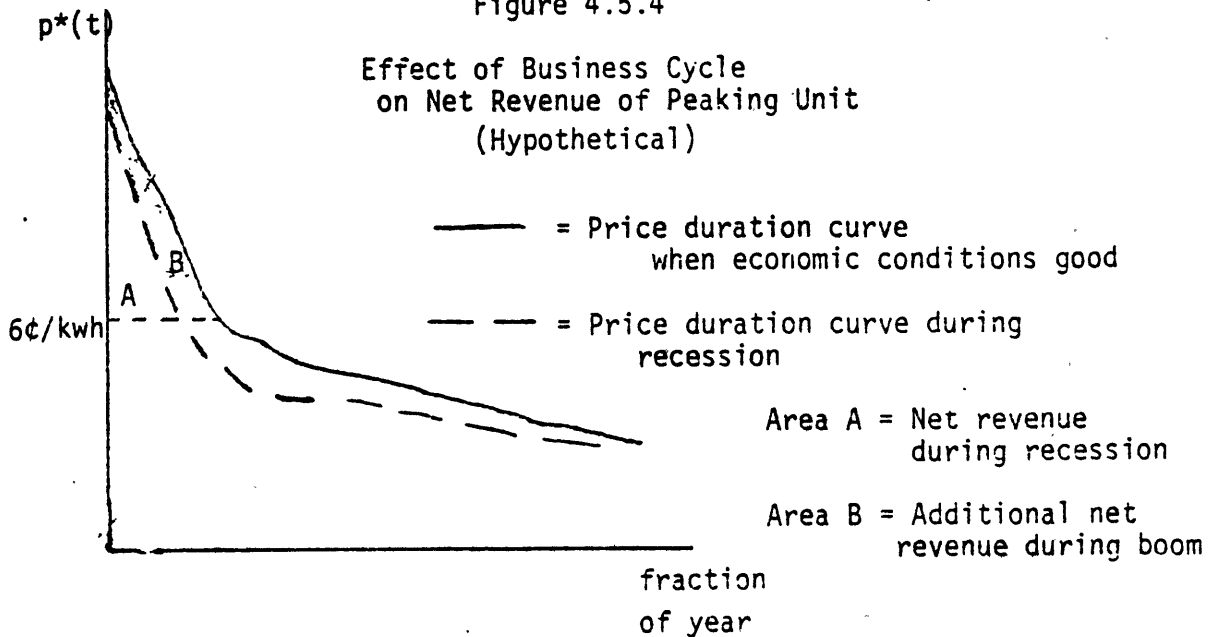


Figure 4.5.4



dependent on the business cycle.^{5.6} But electricity can be stored as an intermediate product only if there is some slack in the production process parameters. When a firm's business is good it may be more profitable for the firm to use this slack to increase total production rather than to store electricity. Therefore demand responsiveness will be inversely correlated with the business cycle, while demand is positively correlated.

This has implications for the riskiness of investments under spot pricing. I will discuss this for an independent generator. The profitability of a generator can be estimated by looking at the area above its marginal operating cost in a price duration curve. Figure 4.5.4 shows such an area. (See Figure 3.1.4 for the derivation.) The above analysis of price responsiveness and the business cycle indicates that the upper left portion of the price duration curve will be quite responsive to the business cycle. Good economic conditions will raise most of the curve and also make it steeper. Therefore investments in peaking plants will have much higher asset beta (undiversifiable risk) than do investments in baseload plants.^{5.7} Therefore in a fully deregulated system or an optimally configured system under full spot prices, reserve margins would be lower than expected by strict expected profit maximization.

Application to Other Commodities

The model of this chapter is applicable to any input to a production process, such as natural gas. Many non-obvious commodities

can be "stored" as intermediate products by end users.

However the interesting time scales for most public utilities are much longer than for electricity, since central storage for short periods is much cheaper and therefore is used to damp daily fluctuations in demand. Thus different response patterns may dominate utilities such as natural gas and water. For example, for natural gas, "shutdown" behavior at the current equivalent prices of fuel oil for each customer's process could set an effective ceiling on natural gas full spot prices.

For telephone services, as for electricity, fast time scales are crucial. An optimal spot price for long-distance telephone communications will have pronounced daily cycles.^{5.8} "Storage" of telephone services for several hours is possible, especially for one-way messages such as data dumps from one computer to another. Of course, the same results apply as for electricity; customers will have higher expected profits under spot prices than under equivalent time-of-use prices, except for some of the customers which are cross-subsidized by time-of-use prices. All customers will have a higher value for demand shifting capital investments under spot pricing.

FOOTNOTES TO CHAPTER 4

1. Many others have developed models of firm behavior under uncertainty. See the survey in Hey [1979]. However these models are oriented toward the effects of risk aversion and contingent claims markets on behavior. They assume very simple production functions to make the analysis simpler. In contrast the model of this chapter looks in detail at the interaction of the firm's production function and price behavior, and assumes risk neutrality and no contingent claims markets. These two approaches are complementary.

2. The full model is a dynamic programming model which would be too expensive to implement in many situations. I therefore discuss computational simplifications which will give reasonable results.

3. The European data are studied in Mitchell et al. [1979]. The same group is beginning to publish analysis of California firms' response to time-of-use prices.

4. Some of the work on time-of-use rates does provide useful insights, since if a firm does not respond at all to them it will not respond to spot price fluctuations of the same or lower magnitude. See especially Reynolds and Creighton [1980].

4. Panzar and Willig [1979] show that under some circumstances symmetry can be used to estimate response to time-of-day electricity prices using data on demand given a time of day pattern of wages. However labor wages do not normally vary hour to hour or on a spot basis, so their method won't solve the problem here.

1.1 Series-parallel processes are also possible. This and many other extensions can be handled within the same basic framework of a dynamic linear or nonlinear programming problem. See Johnson and Montgomery [1974, Chapter 4].

1.2 Production can be measured in any units. It is convenient to normalize so that one unit of final output requires exactly one unit of intermediate product at each stage.

1.3 This is obviously analogous to the use of multiple generating plants to meet total demand, as in Chapter 3. Therefore I will not discuss the implication further.

1.4 It is normal in production planning models to include an inventory carrying cost. These could easily be added.

1.5 Equation 4.1.2a assumes that stage n takes exactly one period. For continuous processes this is not a limitation. For batch processes the equation can be modified.

1.6 A partial curtailment in this model would correspond to a limit on $\sum_n e_n(X_n(t))$. Thus this model can also be used to predict behavior under various rationing methods.

1.7 See Johnson and Montgomery, Section 4-3.2; Tsitsiklas, Section 3.2.

2.1 The model (4.1.1, 4.1.2) might not be strictly accurate, however. Non-convexities such as discrete activity levels would not fit those equations. I will briefly discuss non-convexities later. A model which includes them can be constructed, but might have much longer solution times.

2.2 This ignores another nonlinearity. It may take more energy to pump a gallon when storage is nearly full. Thus e_1 depends on $S_1(t)$.

2.3 Instead of an absolute constraint on storage, costs for over or under heating could be added to the objective function.

2.4 In practice there is an integer programming issue here. This will be covered in Chapter 5.

2.5 In fact, most fossil generators have fixed startup/shutdown costs and constraints. This can be modeled by adding a 0-1 decision variable to the objective function and equation 4.1.4. The consequences of these costs were covered in Bohn [1981]. Basically, the firm will not change its state (operating or shutdown) in response to a momentary change in spot price. It must anticipate that the price will persist long enough to justify the cost of the startup or shutdown.

3.1 For example, it might be optimal for one stage to produce more at a time of high spot prices, in order to make its product available to a more electricity-intensive downstream stage at a later time with lower prices.

3.2 Systems with mandatory final production can still be modeled by setting $|C_2|$ to a very high number, so that it is optimal to always produce $X_{\max,2}$.

3.3 This makes the model unsuitable for thermal storage. Tsitsiklas discusses a clever way to transform the prices to simulate the effects of losses in storage.

3.4 For concreteness I will speak of a cycle length T as one day, and a period length as one hour.

3.5 In fact, full spot prices for most utilities apparently have only one or two significant peaks. Stochastic variation may lead to small additional peaks, but they have small effects on optimal behavior.

3.6 This assumes $L = 1$. If $L < 1$, it may be optimal to wait, even if the price rises.

3.7 Except for special cases, it will always be optimal to set X_n to either its maximum or minimum constraints at each moment. This comes from the assumption that marginal electricity use $[\partial e_n / \partial X_n]$ is constant. See the Appendix.

3.8 When $s_1 = 0$ stages 1 and 2 are functionally equivalent and must be operated in strict synchrony. The pure shutdown case was extensively discussed in Bohn [1981]. If there are fixed costs to shutting down or starting back up, optimal behavior becomes more complex. See footnote 2.5 of this chapter.

3.9 A traditional downward sloping demand curve is the sum of an infinite number of small shutdown processes, each with a different threshold C_2 , and therefore a different option value.

3.10 This can easily be proved by replacing $X_{\max, n}$ with a dummy decision variable in equation (4.1.2b), and adding a single constraint that the dummy must be $\leq X_{\max, n}$.

3.11 The largest possible values for $\sum [a_1(t) + b_1(t)]$ occur when s_1 is less than the interval between price changes. In that case the total of these two shadow prices = total value of the option to store = the difference between minimum and maximum prices each day, in the one peak case. This provides an easy way to estimate the value of small amounts of storage under different rates. I will use this simple case in Chapter 5.

3.12 For example, presently the existence of storage in thermal systems such as air conditioners is essentially based on anticipated demand fluctuations, not price fluctuations. Chillers are sized according to conservative design criteria to meet a "worst day in n years" chilling demand load. As a result, on almost all days chilling capacity $X_{1, \max}$ is greater than chilling load $X_{2, \max}$. Hence air conditioning systems are designed to cycle on and off, allowing temperatures to fluctuate between the set points. These set points then determine the storage and recharge times, s_1 and x_1 in equation (4.3.6). Saving money by taking advantage of price fluctuations was not considered in the design. Under time-varying prices the small amount of existing storage may

therefore have a shadow price much higher than the marginal cost of building more storage capacity. The numerical studies in Chapter 5 will also show that the change from conventional time-of-use rates to full spot pricing may increase these shadow prices quite a bit more.

This example generalizes. Plants today are designed with some capacity in many stages which is "excessive" at most times. These plant designs were optimized considering the following issues:

- o Fluctuations of final product demand
- o Keeping the plant as a whole operating despite partial outages at different stages.
- o Changing final product mix leading to different relative demands in different stages.
- o Fluctuations in labor prices (if labor costs were constant over a week, around-the-clock operation would be the norm, to decrease capital requirements).
- o Occasionally, discrete equipment size availability. (It may be cheaper to buy a 50 hp motor than a 47 hp motor even if 47 hp is adequate.)

Spot pricing or other time-varying prices introduce a new advantage for building "oversized" stages with buffers between them, while not changing any of the earlier incentives. Hence optimal plant design for processes suited to storage or shutdown behavior will change in the direction of more flexibility.

3.13 The value of the first kwh of storage capacity is still the same, since it is given by the peak to trough distance in the price cycle.

3.14 If x_1 were much larger, $\beta^* x_1$ would fall to zero as there would not be enough time to fill storage. Then $T_e - T_d$ would be less than s_1 , and T_f would be less than x_1 . See the appendix.

3.15 This relation is that, unless storage can be fully recharged during the valley, the prices at which discharge begin and end must be at least as high for the second peak as for the first peak.

3.16 I have developed an algorithm to solve the two-peak pure storage case. Because the structure of the problem is so well defined, it is relatively simple. It has not been implemented, but should be quite fast.

3.17 Maintaining the simplifying assumptions used earlier.

3.18 The result that a firm would prefer spot to predetermined prices was derived by Oi [1961]. The issue of "the benefits of price instability" was debated by Samuelson [1972] and Oi [1972]. The essence of their debate in terms of my model is as follows. Since my production model holds equally for a firm which uses or which produces electricity, the analysis appears to imply that artificially induced price fluctuations would improve total profits! However, as we know from Section 3.1, it cannot be that both generators and users are free to adjust their net demand in response to price changes. If markets are to clear at each instant, as they must for an electric power system, then either (a) generators are not free to adjust output in response to price, and Chapter 4's model does not apply to them, or (b) only one (or very few) spot prices are feasible, and they cannot have artificially added price instability. Thus if market-clearing prices happen to be stochastically varying prices it is better not to stabilize them, at least in the zero transactions cost world. But artificially induced instability does not further improve welfare. The purpose of this chapter is to illuminate the critical issue of how fast firms can respond to price changes, since that determines in part whether there is any harm in stabilizing prices.

Nor will artificial demand instability help a firm. Suppose a firm adds a random demand with mean 0 to its existing demand, where those demands are calculated by conventional expected profit maximization. (They may be somewhat stochastic, or deterministic.) Then it will usually demand either more or less than its previous profit-maximizing amount. By the assumption that its previous demand levels were expected profit-maximizing, the new demand level cannot on average make it better off.

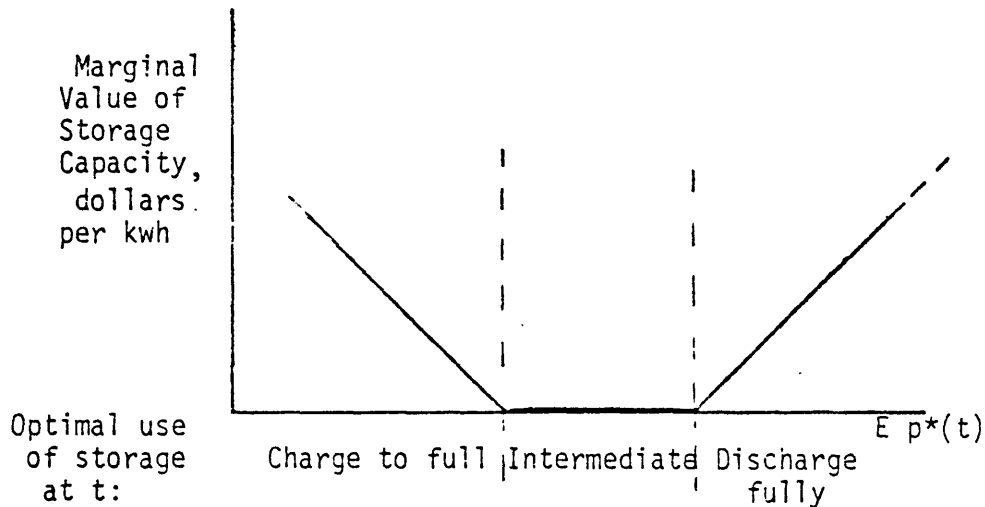
3.19 Whether it is optimal to discharge storage or just hold it level depends on prices after period 7.

3.20 This is not true for the model of this chapter if there are fixed costs to shutdown actions. Knowing the future level will be necessary for optimal behavior by such shutdown customers.

3.21 In fact much simpler information is sufficient in the one or two price peak cases. If there are currently n hours of intermediate product in storage, optimal behavior at t can be determined if the sign of $p(t+n) - p(t)$ is known with certainty. This may be useful for practical forecasting and control under spot pricing. These statements can be rigorously proven using the continuous time model of the appendix.

3.22 The argument is as follows: $b(t)$ is the value of an option to store intermediate product from an earlier time to a later time. As shown in Figure 4.3.8, this means it is convex in expected price, hence its expectation lies on or above its value with predetermined prices.

Figure 4.3.8
Incremental Value of Storage Capacity
as a Function of Anticipated Prices



3.23 Such subsidies are possible if the two rates have prices with different expected values. Customers with correlated demands can have subsidies even if two rates have identical expected values, as discussed in Section 3.5.

3.24 One aspect of this is only conjectured, not proven. See the preceding discussion.

4.1 Utilities have apparently not attempted to exploit the demand smoothing potential of fuel adjustment charges. To do so they would have to publicize the current price. Also the price calculation mechanism should base each month's price on forecasts of fuel charges for that month, more than on costs in past months.

4.2 The following is taken from Acton et al. [1980].

4.3 The equivalent energy charge is impossible to calculate exactly, but I estimate it is about 50 cents/kWh based on the information in Acton et al. The rate is considered to be a "load management" rate rather than "spot pricing;" under either name, it leads to stochastic prices.

5.1 See footnote 3.12 of this chapter.

5.2 Once available stocks of the commodity are empty, more rapid anticipated declines in price are possible.

5.3 Time-of-use prices also encourage storage and damp price cycles. However they encourage less storage (Section 4.3) and have to be set to the "average" daily price cycle, which will only by chance be the day's actual price cycle. Also if left unchanged from year to year they can encourage overinvestment and over-response. This is the "shifting peak" problem.

5.4 Spot pricing will not eliminate anticipated cycles entirely, unless there turn out to be large amounts of potential storage behavior with literally zero marginal storage costs. As price fluctuations are reduced by storage behavior, this reduces the incentives for further storage investments. It also means that firms will find it unprofitable to use thermal storage with high losses, even if they already own the storage equipment. Eventually equilibrium is reached with all the lowest investment and operating cost opportunities taken, but some regular price fluctuation remaining. In contrast, several European countries have instituted time-of-use rates which they left unchanged for many years, leading to a "peak reversal" phenomenon.

5.5 To elicit a full storage response the rise must be anticipated $x_n = S_{\max, n} / (X_{\max, n+1} - X_{\max, n})$ hours in advance. This gives time to fully charge storage. x_n can be anywhere from 10 minutes to a week. However, the height of the rise does not have to be known in advance.

5.6 In a more accurate model the real interest rate, which determines storage costs, is a procyclic variable. Therefore the effect I am discussing here may exist and be empirically verifiable for most commodities.

5.7 Fred Pickel first pointed out to me that different kinds of generation capacity may have different asset betas, even under conventional pricing. Spot pricing will flatten the price duration curve but will also make it more sensitive to the business cycle, strengthening this effect.

5.8 Calculating spot prices for telephones is beyond the scope of this thesis. See Vickrey [1972] for one approach.

APPENDIX TO CHAPTER FOUR
PROOF OF OPTIMAL BEHAVIOR

This appendix proves the optimality of the behavior presented graphically in Section 4.3 for storage and mixed storage/shutdown customers. The approach is to use the first-order conditions to work backwards: If optimal behavior at a particular time is X , what does that imply about the behavior of spot prices. One can then reverse this and determine optimal behavior as a function of spot prices.

Only deterministic prices will be covered. As shown in Section 4.3, customers under spot prices can always use adaptive certainty equivalent behavior, i.e. pretend future prices are deterministic at their expected values. Usually they can do better, but I will not discuss that here.

Model Formulation

To allow easier solution, reformulate the model of (4.3.1) as a continuous-time model. Keep the simplifying assumptions and definitions:

$-C_2$ = value of final product = constant

e_1 = constant = 1 for simplicity

L = 1 - losses in storage

$S_1(0)$ = 0 = initial storage.

Also $X_{\max,1} > X_{\max,2}$. Otherwise the firm could sell everything it produced, and storage would be irrelevant.

The Hamiltonian for the cost minimization problem is then:

$$\begin{aligned}
 (4.A.1) \quad H(t) = & P_1 X_1(t) + C_2 X_2(t) \\
 & + \alpha_1(t)[X_1(t) - X_{\max,1}] - \dot{\alpha}_1(t)X_1(t) \\
 & + \alpha_2(t)[X_2(t) - X_{\max,2}(t)] - \dot{\alpha}_2(t)X_2(t) \\
 & + \beta_1(t)[S_1(t) - S_{\max,1}] - \dot{\beta}_1(t)S_1(t) \\
 & - \gamma(t)[X_1(t) - X_2(t)]
 \end{aligned}$$

where $\dot{S}_1 = dS_1/dt$ and $P_1(t)$ is the per unit variable cost for stage 1 at time t , including the cost of electricity. (4.A.1 is the continuous time counterpart of the LP discussed in Section 4.3.) Other variables are as in Section 4.3, except that $\dot{\alpha}_1$, $\dot{\alpha}_2$, $\dot{\beta}_1$ are non-positive dual variables on the non-negativity constraints, and γ is minus the costate variable corresponding to the state $S_1(t)$. As defined here γ will be positive.

First-order conditions include:

$$\begin{aligned}
 \partial H / \partial X_1 &= P_1(t) + \alpha_1(t) - \dot{\alpha}_1(t) - \gamma(t) = 0 \\
 \partial H / \partial X_2 &= -C_2 + \alpha_2(t) - \dot{\alpha}_2(t) + \gamma(t) = 0 \\
 \dot{\gamma} &= \partial H / \partial S_1 = \beta_1(t) - \dot{\beta}_1(t)
 \end{aligned}$$

Complementary slackness holds:

$$\begin{aligned}
 0 &= \alpha_1(t)[X_1(t) - X_{\max,1}] = \dot{\alpha}_1(t)[X_1(t) - 0] \\
 &= [\alpha_1(t)] [\dot{\alpha}_1(t)] \quad \text{etc.}
 \end{aligned}$$

There will also be initial conditions ($S_1(0) = 0$ for example) and terminal conditions, which I do not use.

I will assume price $P_1(t)$ changes continuously, i.e. $\dot{P}_1(t)$ always exists and is only momentarily zero. The costate variable $\gamma(t)$ will also be continuous. I will drop the time arguments for brevity. Remember $C_2 < 0$, so $|C_2| = -C_2 > 0$. $|C_2|$ is the value of each unit of final product, X_2 .

Note that the Hamiltonian is linear in X_1 and X_2 . Hence the optimal solution will always be at an extreme point (unless it is degenerate). Because of this linearity, second order conditions cannot be used to distinguish maxima from minima. Instead, we can use inspection. Since this is a cost minimization problem, proper behavior is to use the least electricity at times of high prices, and the most at times of low prices. (See the figures at the end of the appendix for illustrations.) Specifically:

- o Since $C_2 = \text{constant}$, it never pays to defer final production X_2 to a later time. So X_2 will always be at its physical maximum:

$$\begin{aligned} X_2 &= \text{Min} [X_1; X_{\max,2}^n] \text{ when } S_1 = 0. \\ (4.A.2) \quad &= X_{\max,2} \text{ when } S_1 > 0. \end{aligned}$$

- o If prices are always changing, then $0 < X_1 < X_{\max,1}$ is always dominated by $X_1 = 0$ or $X_1 = X_{\max,1}$. Even if $P_1 = 0$ over some interval, an extreme value of X_1 is optimal although it may not be unique.

Possible States

At any instant storage is either full, empty, charging, or discharging, or changing from one state to another. In each case we

can deduce what is happening to prices at that instant. The key will be the costate variable, γ . This will serve to prove the optimality of the diagrammatic solutions in the main text.

Case 1. Charging: $\dot{S}_1 > 0$, which can only happen if $X_1 > X_2$.

This implies

$$\begin{aligned}\gamma &= P_1 + \alpha_1 \\ &= C_2 - \alpha_2 + \alpha_2 \\ \beta_1 &= 0 = \dot{\gamma} \text{ (by complementary slackness on } S_1 \text{)}.\end{aligned}$$

By the argument that X_1 and X_2 will always be extreme values, $X_1 = X_{\max,1}$; $X_2 = X_{\max,2}$. This implies

$$\begin{aligned}\gamma &= P_1 + \alpha_1 \text{ and } \gamma = C_2 - \alpha_2 \\ \text{Since } \dot{\gamma} &= 0, \dot{\alpha}_1 = -\dot{P}_1\end{aligned}$$

Spot price P_1 may be rising or falling.

Case 2. Holding at full: $\dot{S}_1 = 0$ so $X_1 = X_2$

$$S_1 = S_{\max,1} \quad \text{so } \beta_1 > 0$$

Case a. Maximum production: $X_1 = X_2 = X_{\max,2}$ so

$$\begin{aligned}\alpha_1 &= 0, \alpha_2 > 0 \\ \gamma &= C_2 - \alpha_2 \\ \dot{P}_1 &= \dot{\gamma} = \beta_1 > 0. \text{ Spot prices must be } \underline{\text{rising}}.\end{aligned}$$

Case b. $X_1 = X_2 = 0$.

As discussed earlier, when storage is full it will never be a unique optimal solution to a set $X_2 = 0$. Therefore this case will never be better than Case a, and can be ignored without loss of generality.

Case 3. Discharging: $S_1 < 0$ so $X_1 < X_{\max,1}$ and $X_2 > 0$.

By the earlier argument this implies $X_1 = 0$, and $X_2 = X_{\max,2}$.

$$\dot{\gamma} = P_1 - \alpha_1$$

$$\dot{\gamma} = |C_2| - \alpha_2$$

$$\dot{\gamma} = \beta_1 = 0$$

Prices can be rising or falling. But, discharge can only occur after charging. (Since started with $S_1(0) = 0$.) So prices must at some previous time have been rising.

Case 4: Holding at empty: $S_1 = 0$; $\dot{S}_1 = 0$; $X_1 = X_2$.

Case a. $X_1 = X_2 = X_{\max,2}$

$$\dot{\gamma} = P_1 = |C_2| - \alpha_2$$

$$-\beta_1 = \dot{\gamma} < 0.$$

So prices must be falling.

Case b. $X_1 = X_2 = 0$.

$$\dot{\gamma} = P_1 - \alpha_1 = |C_2| + \alpha_2$$

So

$$|C_2| = P_1 - \alpha_1 - \alpha_2 < P_1$$

This is the shutdown state: $X_1 = X_2 = 0$. It is only optimal if

$|C_2| < P_1$ and $S_1 = 0$. In this situation, extra production costs more than the product can be sold for. See Figure 4.3.5.

Transitions

Now consider the possible transitions from one state to another. The possible transitions are discussed in Cases 5 through 8, as

<u>From:</u>	<u>T₀</u>			
	<u>Charging</u>	<u>Full</u>	<u>Discharging</u>	<u>Empty</u>
<u>Charging</u>	X	Case 8	Case 5	X
<u>Full</u>	X	X	Case 5	X
<u>Discharging</u>	Case 7	X	X	Case 6
<u>Empty</u>	Case 7	X	X	X

Case 5. Transition from another state to discharging. Prices at some previous time were rising, during which storage was filled, partly or fully. The possible earlier states are charging (Case 1) or holding at full (Case 2). Call the instant at which discharge begins T_d .

From Case 3,

$$r(T_d) = P_1(T_d) - \alpha_1(T_d)$$

But from Case 1 or Case 2,

$$r(T_d) = P_1(T_d) - \alpha_1(T_d)$$

By continuity of $\gamma(t)$, this means

$$\begin{aligned}\gamma(T_d) &= P_1(T_d) \\ \alpha_1'(T_d) &= \alpha_1(T_d) = 0.\end{aligned}$$

So for the entire period of discharging,

$$\gamma(t) = P_1(T_d). \quad (\text{See Case 3 for proof } \dot{\gamma} = 0.)$$

Case 6. Transition from discharging to holding at empty. (This is the only possible transition to empty.) Call the time of transition T_e .

Case a.: $|C_2| > P_1(T_e)$, so keep producing.

$$\gamma(T_e) = P_1(T_e). \quad \text{This implies } P_1(T_d) = P_1(T_e)$$

This is a key claim in Chapter 4: Unless spot prices rise so high as to make a shutdown optimal, the spot price at the time storage optimally begins to empty will equal the spot price when it stops emptying.

Case b.: If $|C_2| < P_1(T_e)$, Case 4b applies.

$$\gamma(T_e) = P_1(T_e) - \alpha_1'$$

So $0 \leq \alpha_1'(T_e) = P_1(T_e) - P_1(T_d)$

Prices must have stayed the same or risen since storage discharging was begun. The firm will remain shut down until the spot price falls below $|C_2|$.

Case 7: Transition from another state to charging. The preceding state can be either holding on empty (e.g., at the beginning of the day), or directly from discharging. This applies to the two-peak cases, when there is not enough time between peaks to completely refill storage. Call the time of transition T_c .

From Case 1 (charging) we have:

$$\gamma(T_c) = P_1(T_c) + \alpha_1(T_c)$$

From Cases 3 or 4a:

$$\gamma(T_c) = P_1(T_c) - \alpha_1(T_c)$$

so by continuity,

$$\gamma(T_c) = P_1(T_c)$$

$$\alpha_1(T_c) = \alpha_1(T_c) = 0$$

Case 8: Transition from charging to full. Let the time at which this occurs be T_f .

$$\text{From Case 1 we have: } \gamma(T_f) = P_1(T_f) + \alpha_1(T_f)$$

$$\text{From Case 2 we have: } \gamma(T_f) = P_1(T_f)$$

$$\text{so } \alpha_1(T_f) = 0$$

Again, we know $0 = \dot{\gamma}$ throughout charging, so

$$P_1(T_c) = \gamma(T_c) = \gamma(T_f) = P_1(T_f)$$

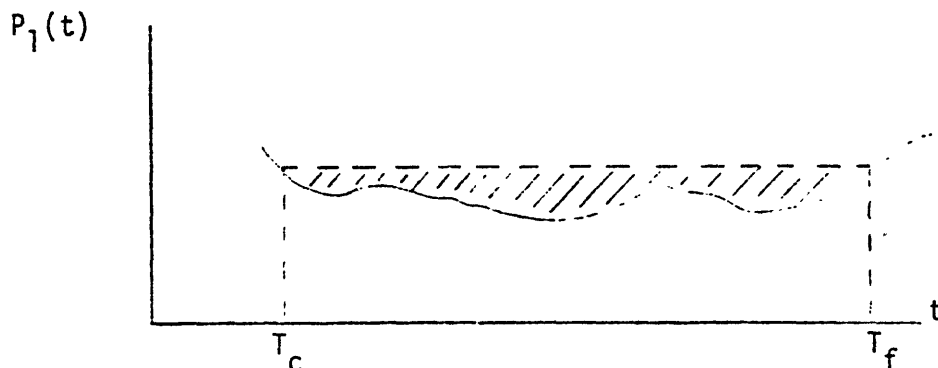
The spot price when charging begins and ends must be equal. From case 2, we know that this price must be at or below the spot price at the time discharging next begins.

Putting the Cases Together

We can now derive the properties of the complete optimal solution over time. A charging episode, which begins at time T_c and ends at time T_f , must have those times such that price P_1 is equal at both times. This price includes the electricity price plus any prices of other variable inputs (such as labor). By the conditions for cost minimization, the price between T_c and T_f must be lower than the price at either end. This implies $\dot{P}_1(T_c) < 0$ and $\dot{P}_1(T_f) > 0$. Therefore there must exist at least one t , $T_c < t < T_f$ such that $\dot{P}_1(t) = 0$. An example of a possible charging episode is shown in Figure 4.A.1.

Figure 4.A.1

Example of Optimal Charging Episode



Analogous conditions hold for each discharging episode, which begins at T_d and ends at T_e . Price between these times must be higher than at either end, and equal at the ends:

$P_1(T_d) = P_1(T_e)$. Discharging will be done until storage is completely empty, unless charging begins immediately in anticipation of a later peak. That is, it is never optimal to "hold partially empty" after a discharge episode. (Unless price is flat and the solution non-unique.) It will be optimal to "hold on empty" under either of two conditions. First, if $P_1(T_e)$ is higher than $|C_2|$ it is optimal to leave the entire process shut down until price falls below the shutdown threshold. Second, if price continues to fall, it may be optimal to wait before beginning recharging. Figure 4.A.2 shows an example of a discharging episode.

Periods of charging and discharging may lead directly into one another, if major price peaks and valleys are close together. When there are long periods of monotonic price change, it may be optimal to "hold on full" after charging, and "hold on empty" after discharging fully. Figure 4.A.3 shows an example of optimal behavior over two days, with some behavior of each type.

Figure 4.A.2
Example of Optimal Discharging Episode

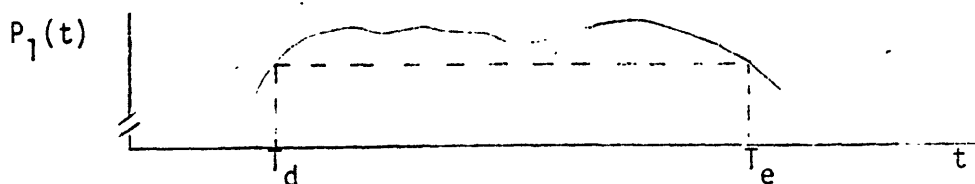
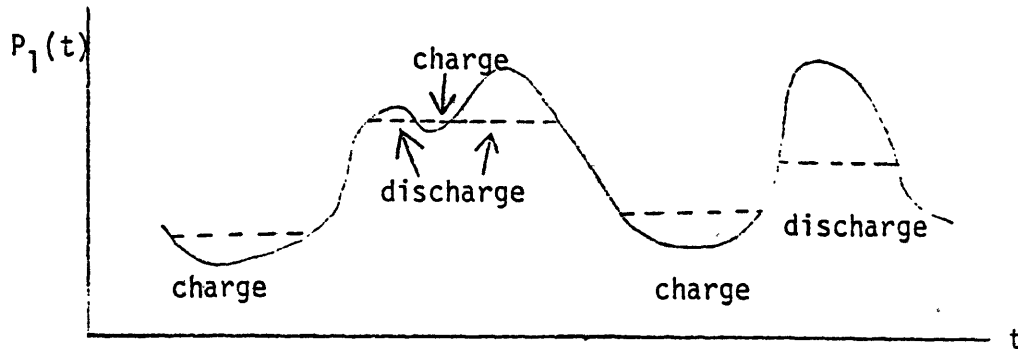


Figure 4.A.3
Example of Optimal Behavior



Investment

The analysis of each case also proves the contention of Chapter 4 that the shadow value of storage capacity is the difference between the prices when charging ends and when discharging begins, summed over all intervals of rising prices. This value is measured in the same units as price, i.e. in dollars per kilowatt hour of storage capacity. Recall that the shadow value of storage is $\beta_1(t)dt$ over the life of the capacity. From Cases 1 through 4 I showed that $\beta_1 > 0$ if and only if "holding on full," and this occurs only when prices are rising. All other states contribute nothing.

Similarly we can now prove that the shadow value of charging capacity, $\alpha_1(t)dt$, is given by the difference between the price at which charging begins and ends, $P_1(T_c)$, and the current price, integrated over each charging episode. This is measured in dollars per kw of charging capacity.

Chapter 5

NET BENEFITS OF SPOT PRICING: A CASE STUDY

This chapter uses case studies to give quantitative estimates of the effects of spot pricing compared with a variety of other rates. In the absence of transactions costs, spot pricing will give results which are socially preferable to those under any other rates. But only if the gross benefits of spot pricing are larger than its additional transactions costs should it be implemented. This chapter shows by example how to make the appropriate calculations. For selected hypothetical customers and one utility, it discusses:

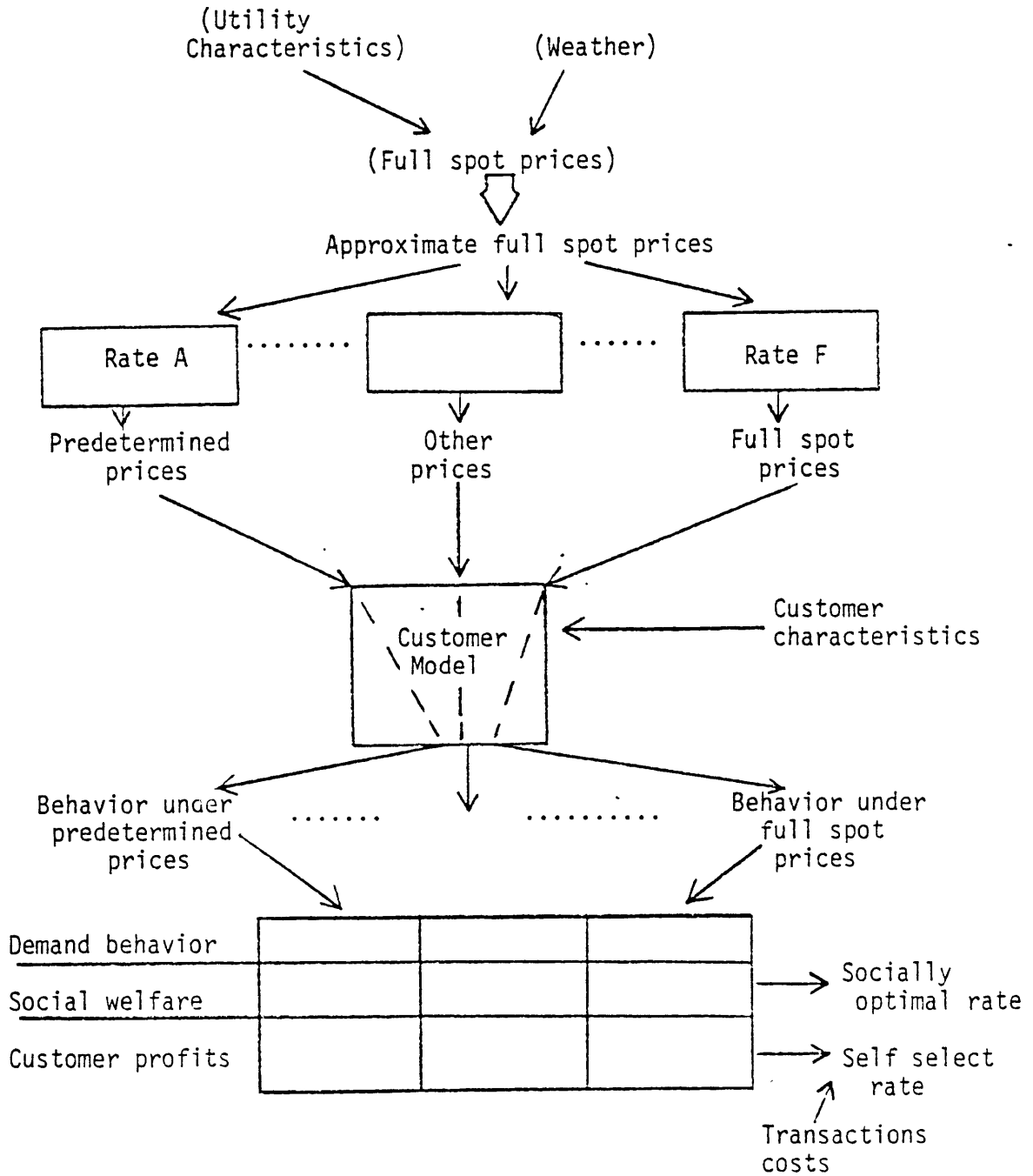
- o How much better is full spot pricing than flat pricing? Will other rates of intermediate sophistication, give most of the benefits of full spot prices?
- o Which rate is socially optimal for each customer?
- o Which rate will each customer select if allowed a free choice?

Figure 5.0.1 illustrates the basic flow of calculations for each case study customer.

Six rates are modeled. Each rate is a different rule of calculating prices.

- o Prices recalculated once a year
 - Flat prices for entire year (Rate A)
 - Prices changed twice a day: (two level time-of-use price) (Rate B)

Figure 5.0.1
Case Study Calculation Flow



- o Prices recalculated once a month
 - Flat prices for entire month (Rate C)
 - Prices changed twice a day (Rate D)
 - Prices changed every hour according to that month's pattern. (Rate E)

- o Prices recalculated each hour
 - Full spot prices; change each hour (Rate F)

For each of these rates, I simulate how prices would have been set under that rate for the test year, 1980. Some rates have considerable time aggregation error and forecasting error and therefore lead to prices which deviate considerably from full spot prices.

Each hypothetical customer is modeled using the mathematical techniques of Chapter 4. This gives an hour-by-hour simulation of how the customer would have behaved under each rate. Behavior under different rates is compared to give:

- o The gross social welfare effect of each rate.
- o The social welfare impact of each rate, net of transactions costs for that rate.
- o The socially optimal rate to which customers of that type should be assigned, as a function of customer size.
- o The analogous results for private profits of the customer.
- o The effects of different rates on the incentives to make various investments.

By comparing results across customers and across different

utilities, we gain additional insights into what determines the relative value of full spot pricing. Key determinants include:

- o the variability and predictability of full spot prices for the utility
- o the customer's size (in MW of maximum demand)
- o the nature of the customer's electricity use.

Several unexpected patterns show up in the results.

The chapter analyzes individual customers and a single utility. Section 5.1 sketches the utility, its full spot prices for 1980, and the various rates which are compared. Section 5.2 describes the customers modeled. Section 5.3 presents the basic results. These include the gross social welfare and gross profits of each customer under each rate, before allowing for transactions costs. Investment incentives under different rates are also shown. Section 5.4 discusses transactions costs and their effects on optimal rules for assigning customers to rates. For each customer type, optimal assignment is a function of customer size.

Section 5.5 presents sensitivity analysis on utility characteristics. Section 5.6 examines the distribution of actual customers across the U.S., using the limited data available for this purpose. It discusses how much electricity is used by those classes of customers which appear to be suited to full or partial spot pricing. Section 5.7 summarizes the results and limitations of the case studies.

5.1 The Utility and Rates Modeled

Detailed data on a single utility in the East North Central region of the U.S. was used for all case studies, and is described in this section. 1980 was used as the case study year. In order to compare the effects of different rates, we need to know what prices would have been under each rate, if it had been in effect. Estimating this requires a series of decisions, which I discuss in this section.

These include:

- o Which rates to use. All the rates I used were based on full spot prices, with different degrees of time aggregation and prespecification.
- o How to estimate full spot prices.
- o How to simulate prices under rates other than full spot pricing. This mainly requires a means of simulating forecasts of future spot prices. Since we have historical data for 1980, we know what the "correct" forecasts were; but actual price setters would not have had access to this information.

I will discuss full spot prices first, to remind readers of their significance. For this utility, they are highly variable.

Estimating Full Spot Prices

The "raw materials" for prices under each of the six rates modeled are the full spot prices for the study year, 1980. From Chapter 3, we have that the full spot prices for customer j at the t is:

$$(3.2.1) \quad p_j^*(t) = \left[1 + \frac{\partial L(t)}{\partial D_j(t)} \right] \left[\lambda(t) + \mu(t) \right] + \text{terms involving T \& D constraints}$$

where

$\lambda(t)$ = System lambda (Short run marginal fuel and operating costs for generation)

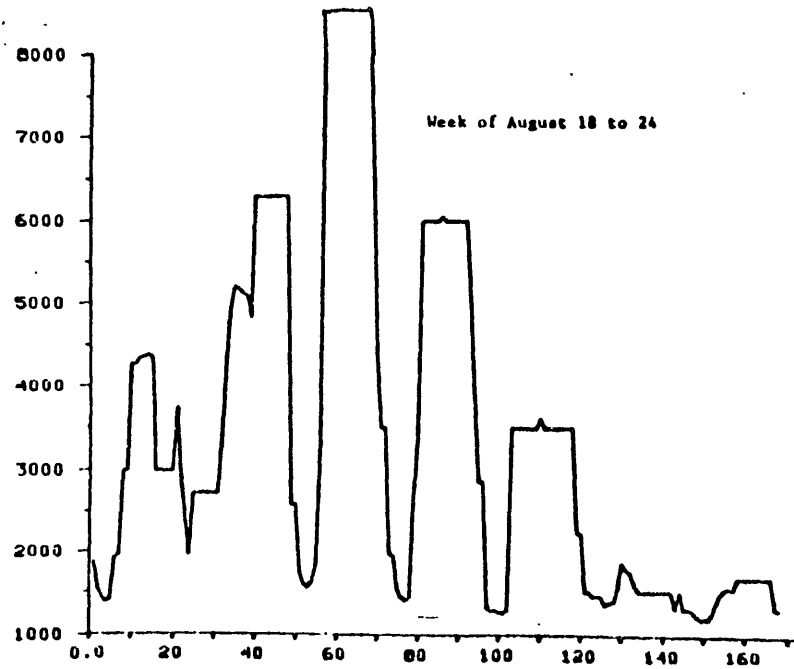
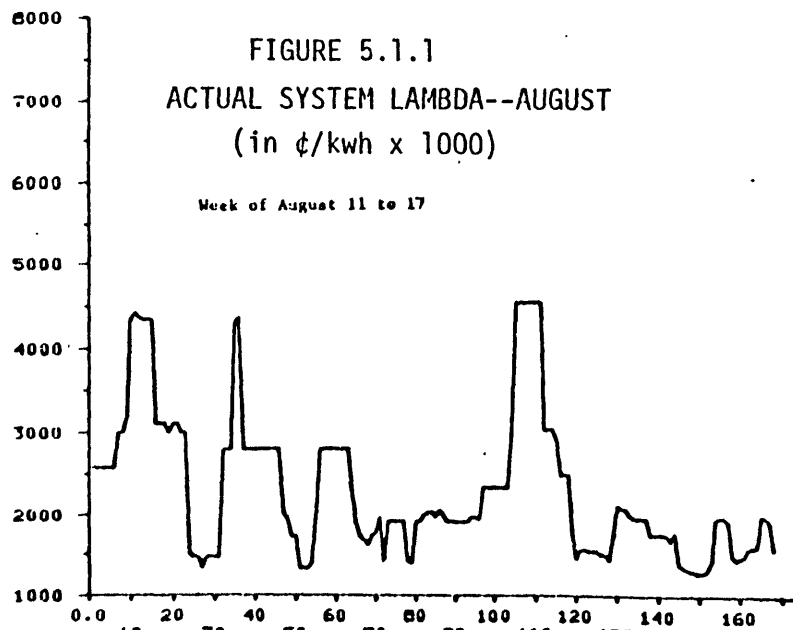
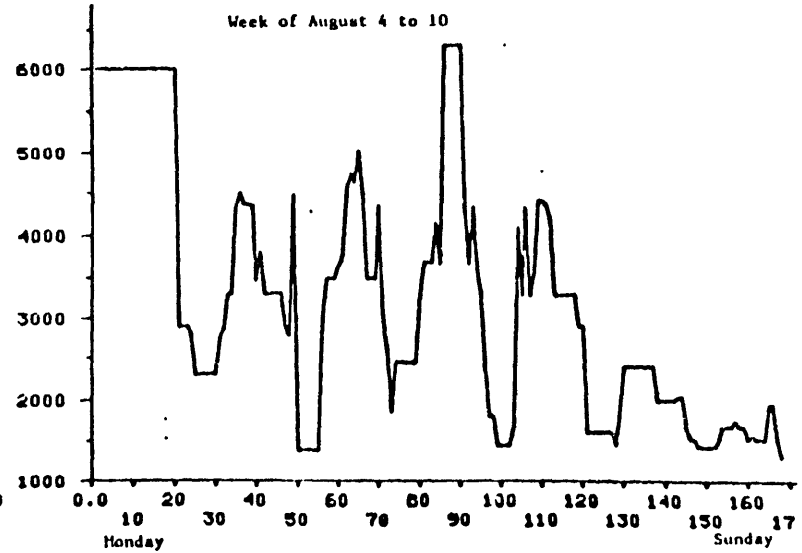
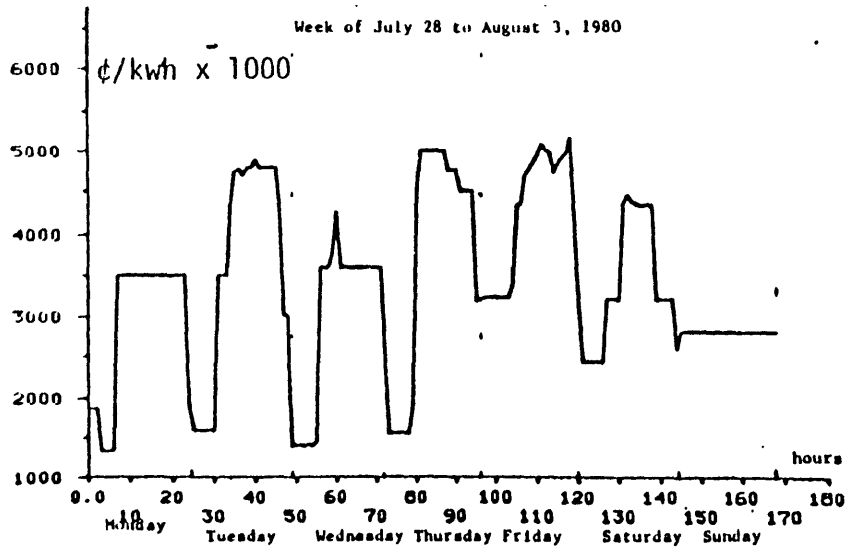
$\mu(t)$ = Curtailment premium

$\partial L / \partial D_j$ = Marginal effect of customer j on losses

Historical system lambda for the subject utility is available from July 1979 to December 31, 1980, hour by hour.^{1.1,1.2} The other components of equation 3.2.1 were not available, and I therefore used the approximation that full spot price equals system lambda.^{1.3}

Chronological plots of system lambda show tremendous variation from week to week, and considerable variation from day to day. Figure 5.1.1 shows four weeks in August, on a consistent scale. The vertical axis is in cents per kilowatt hour; the horizontal axis is hours from 00:00 Monday. The low each night is consistently between 1.0 and 1.5¢/kWh, except for one 4-day period. It almost always comes between midnight and 6 AM. But the daily peaks show considerable variation in amplitude and pattern. Some days show "needle peaks"; others are approximately flat for 12 hours. Thus socially optimal customer behavior will vary considerably from day to day. Figure 5.1.2 shows the second weeks of January, April, October, and December; the variability within August is not extreme compared with other months. Figure 5.1.3 shows the price duration curve for 1980, and summarizes system lambda over the year.^{1.4}

Notice that this utility is an "interesting" candidate for spot



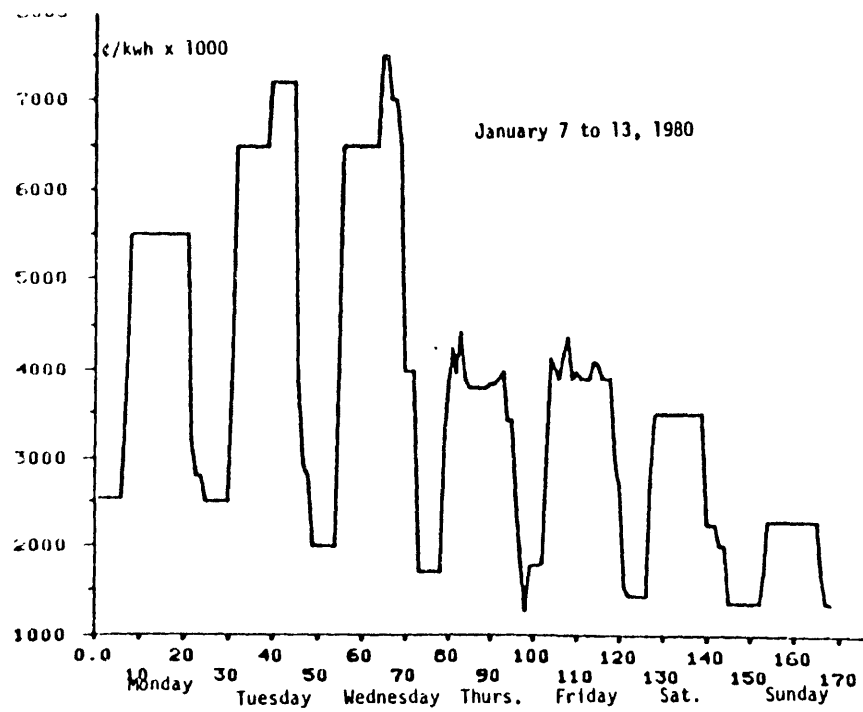


FIGURE 5.1.2
Actual System Lambda--Selected Weeks
(in ¢/kwh x 1000)

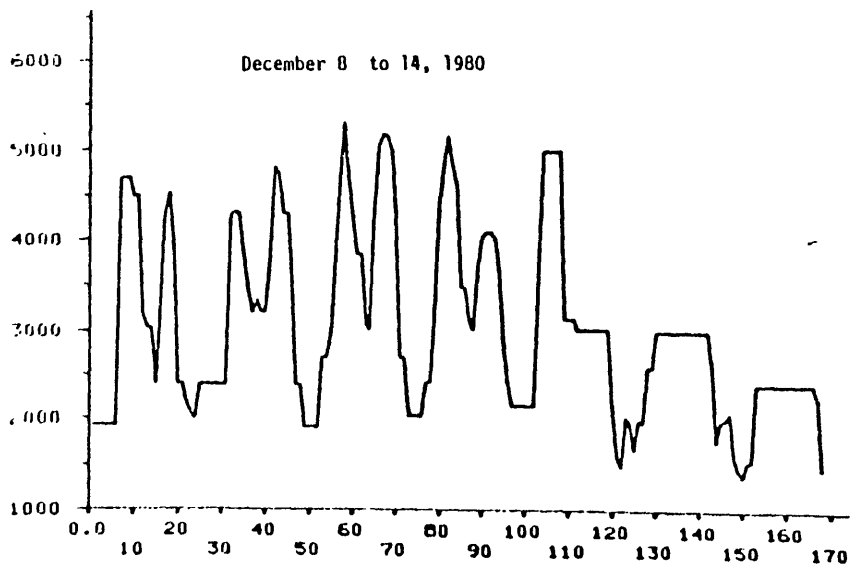
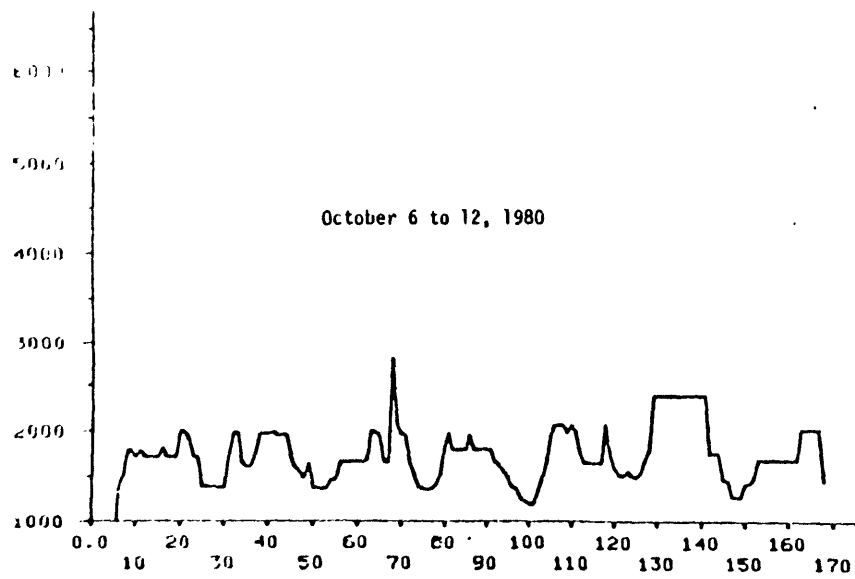
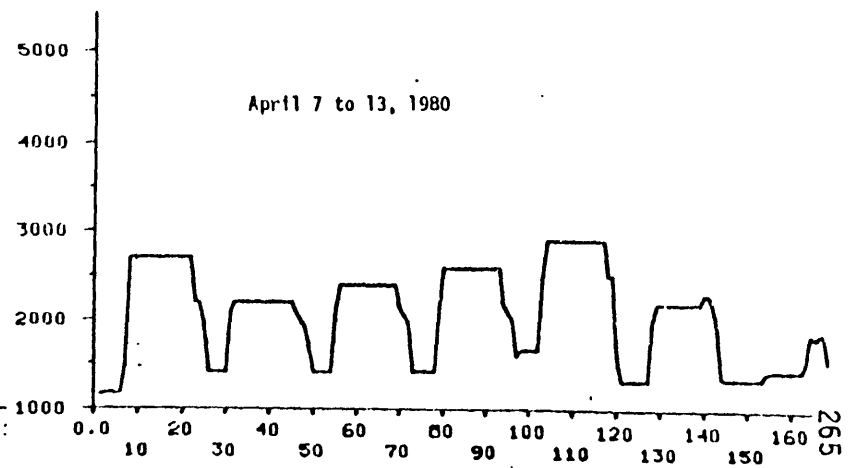
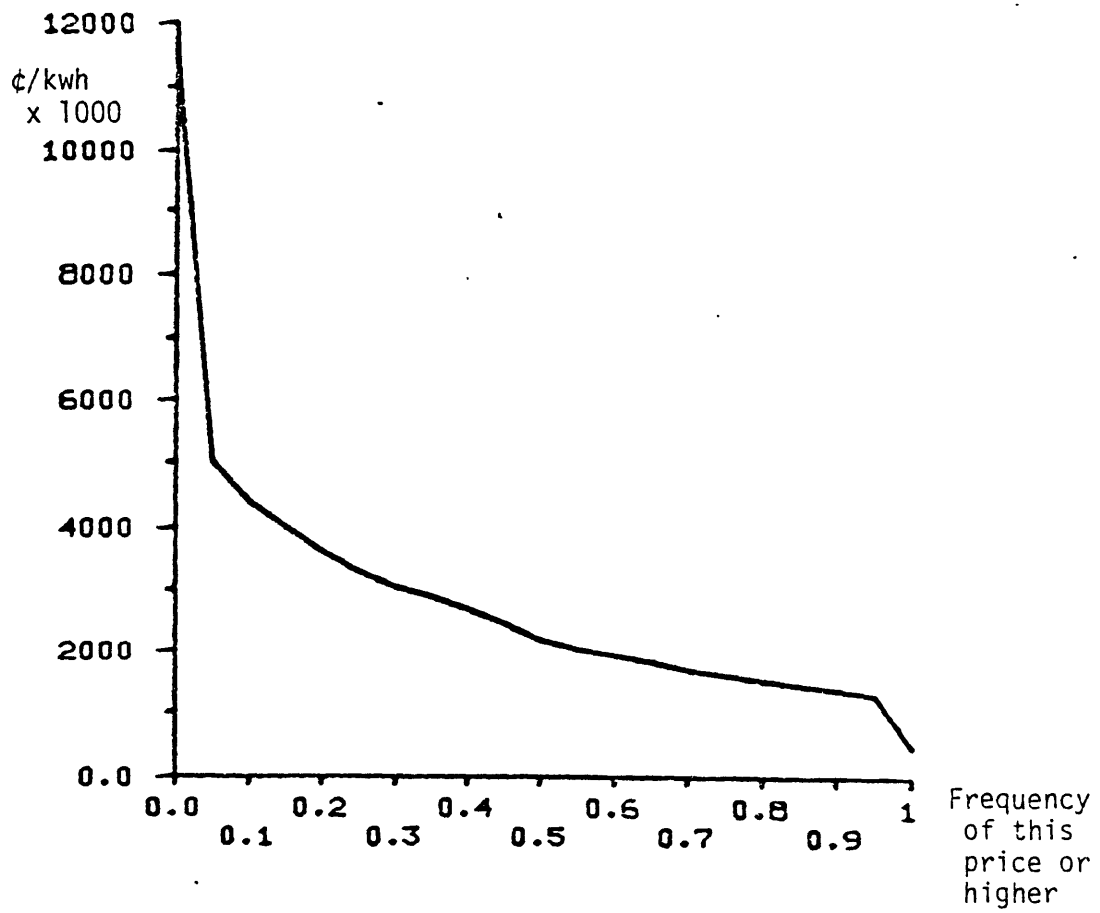


Figure 5.1.3
Price Duration Curve for 1980



pricing, in that its spot price is quite variable and will therefore be poorly approximated by conventional rates. This variability stems from the utility having a mix of generating technologies, from nuclear to gas turbines. Demand swings and unit outages are broad enough that any of these may be the marginal unit within a single week. In contrast some utilities in other regions of the country almost always have a single type of unit on the margin. Such utilities would not show much benefit from spot pricing. I will return to this issue in Section 5.5.

Rates Selected

The rates used are all based on full spot prices, since Section 3.3 showed that optimal non-spot prices are weighted averages of the full spot price at the same time. Thus all six rates are based on consistent principles. In conventional terminology, they would all be considered short run marginal cost based rates. The six were chosen to give a wide range of time aggregation and predetermination levels. In this way we can see if it is necessary to go all the way to full spot prices, or whether some rate with lower transactions costs and lower sophistication can do almost as well.

Figure 5.1.4 shows the six rates. Rate A is fully predetermined and completely flat, while rate F is full spot pricing. Rates A through F are progressively more sophisticated, i.e. closer to full spot prices, and we therefore expect them to show progressively higher gross social benefits.^{1.5} The utility's actual pricing for large

Figure 5.1.4

Rates Used for Case Studies

	<u>Frequency of Price Recalc.</u>	<u>Frequency of Price Change</u>	<u>Forecast Error</u>	<u>Time Aggregation Error</u>
Rate A (Annual flat)	Annual	Annual	High	High
Rate B (Annual TOU)	Annual	Every 12 hours*	High	Medium
Rate C (Monthly flat)	Monthly	Monthly	Medium	High
Rate D (Monthly TOU)	Monthly	Every 12 hours*	Medium	Medium
Rate E (Monthly 24)	Monthly	Hourly	Medium	Zero
Rate F (Full spot)	Hourly	Hourly	Zero	Zero

*Except on weekends

A: Flat price, set at end of 1979 for all of 1980.

B: Two-level time-of-use price, set at end of 1979 for all 1980. The off-peak price covers 8 PM to 8 AM each weekday, and all day on weekends.

C: Flat price, set at end of each month for following month.

D: Two-level time-of-use price, set at end of each month, for following month.

E: Hourly time-of-use price, set at end of each month for following month. In this rate, 24 separate prices are set for weekdays, and 24 for weekends. For example, all Saturdays and Sundays within the month have a single price from 8 AM to 9AM, and another price from 9 AM to 10 AM.

F: Full spot price. A different price is set at the beginning of each hour.

commercial and industrial customers is close to rate B. It is a two level time-of-use price, changed once a year or less often. The peak period is 8 AM to 8 PM, Monday to Friday.^{1.6} New prices or changes in existing prices must be approved by the state regulatory commission.

Prices Under Non-Spot Rates

The next task is to simulate how prices would have been set under each rate, if it had been in effect during 1980. For example under rate A, a single price would have been chosen at the end of 1979 to cover all of 1980. From Section 3.3, we know that the optimal level of this price is approximately

$$\sum_{t=1}^{8784} E_{1979} p^*(t)$$

where $p^*(t)$ is the full spot price during hour t of 1980, and E_{1979} is the expected value operator, conditional on all information available at the end of 1979.^{1.7} Thus setting non-spot prices requires forecasts of spot prices. For this case study we know the true values of 1980 spot prices, but must simulate how they would have been forecast at the end of 1979. I will discuss forecasting below.

A conceptually important set of adjustments to the prices under different rates is to make all of them give the same utility net revenue. This is necessary for regulatory acceptance of the rates, as well as to control the incentives for utility opportunism, as discussed in Section 3.6. However, these adjustments require more data than is available for the case study. Fortunately, it turns out

that they would have relatively little impact on the results. Here I will discuss how the adjustments should be made, given sufficient data; at the relevant point in Section 5.3 I will show why they would have little effect.

Adjusting rates to give the same utility net revenues is a two step procedure. First, the effect of unadjusted rates on net revenues must be calculated. If we use net revenues under full spot prices as the base level, then the impact of customer j on utility net revenues is minus its cross subsidy S_j under the other rate, which was defined in Section 3.5.^{1.8} These cross subsidies must be summed over all customers on the rate, to find the total revenue to be made up. Second, prices are adjusted to counterbalance this lump sum. For example, if net revenues under rate A would be lower than under full spot prices, prices under rate A should be raised. This can be done in a variety of ways; see Section 3.6.

The reason such adjustments cannot be done for this chapter is that the first step requires estimating the effects of an entire customer class on net revenues. The chapter models only individual and hypothetical customers.

Simulating Forecasts of Spot Prices

Each non-spot rate requires the utility to forecast future system lambda, either one year ahead (for rates A and B) or one month ahead (for rates C, D, and E). I considered several methods of simulating this process. All of them have the drawback that the utility had access to forecasts of relevant causal variables which I have little data on. Thus I cannot directly reproduce how the utility would have made forecasts. These variables include:

- o Planned outages of units, e.g. for nuclear plant refueling.
- o Demand and line loss forecasts.
- o Contractually specified fuel prices.
- o Contractually specified sales to or purchases from neighboring utilities.

On the other hand, I know how the interaction of these variables "turned out", since I know the actual system lambdas during 1980.

With these factors in mind I considered the following forecasting methods:

- o Time series forecasts of future system lambda based on past values. Such forecasts would be less accurate than actual utility forecasts, because the utility would have the additional information on other causal values. More important, I only had historical data starting in July 1979, which is less than a full year before the start of the forecast period. This would be completely inadequate for

ARIMA or other time series methods, since the spot prices probably have an annual cycle.

- o Using a month-by-month forecast made by the utility in January 1980, covering 1980 to 1985. The forecast for 1980 was well below the actual level of system lambda in 1980, averaging 15 percent too low during off peak hours. These errors seemed extreme, and would have made predetermined prices perform quite poorly in the case studies. Therefore I did not use these forecasts.
- o Assuming all forecasts were completely accurate, i.e. using the actual 1980 values in place of forecasts. This would overstate the performance of predetermined prices.

For the one year ahead forecasts (rates A and B), I simply exogenously set a forecast error of 5 percent. Thus I used "forecast" system lambdas of 95 percent of the true values. A modified version of this approach would be to "sample" the forecast error from some underlying distribution, and run the model with rates A and B many times, each with a different forecast error. Again, not enough information was available to construct such a probability distribution; the single point estimate used (5 percent) is likely to be near the center of the probability distribution, judging by the much larger errors in the utility's own forecasts, cited above.^{1.9}

For the one month ahead forecasts (needed in Rates C, D, and E) I used a weighted average of the previous month's full spot prices and the actual full spot prices for the month in question. The utility

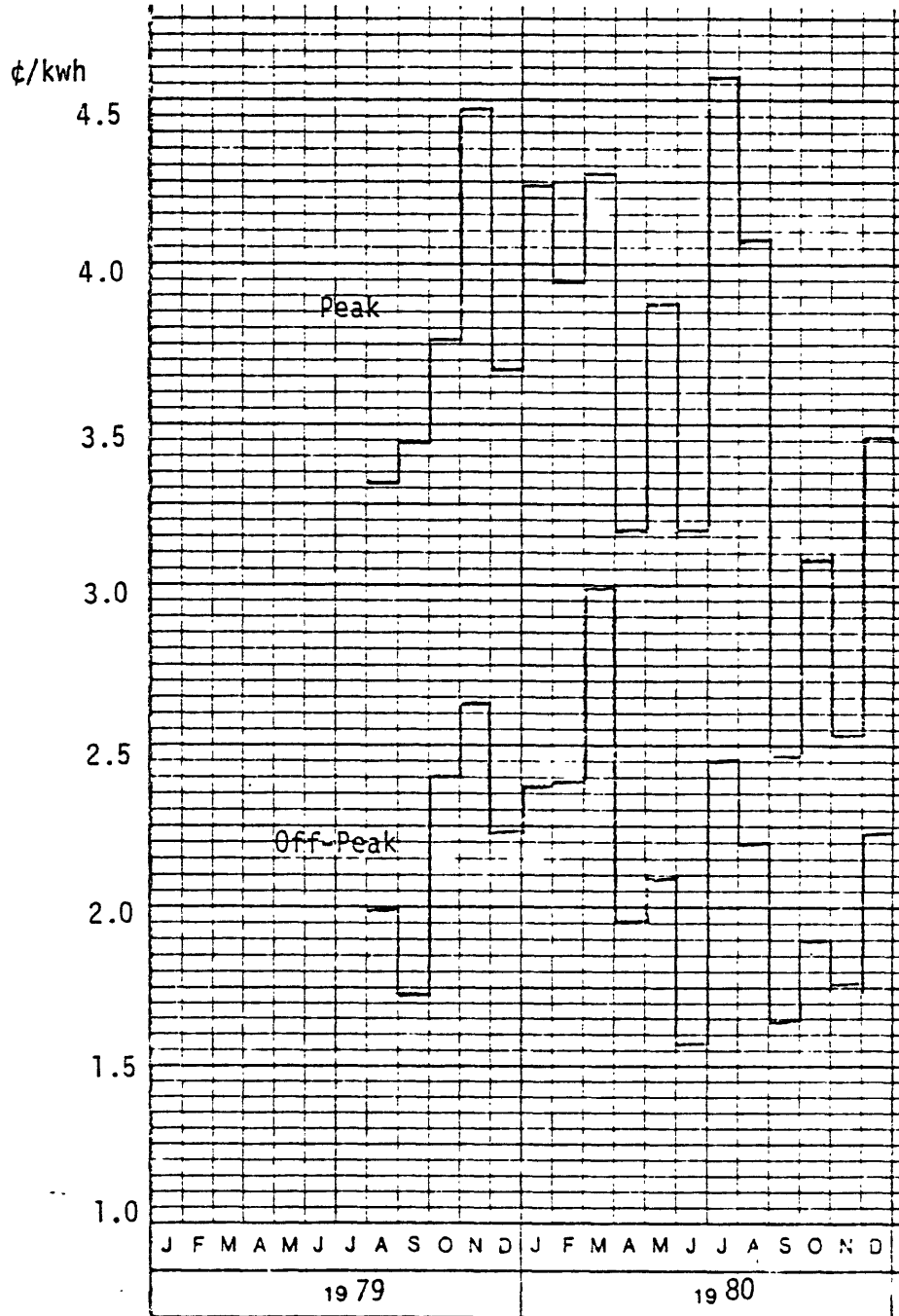
could always use the previous month's spot prices with a weight of one; a lower weight on historical information corresponds to better knowledge of the charges which will take place the next month. I used a weight of .6 on the past month and .4 on the actual month. Figure 5.1.5 shows the actual system lambdas from month to month, for both peak and off-peak periods. Note the considerable month to month variation.

The weights of .6/.4 imply that slightly more than half of the month-to-month change in system lambda was not predictable. To determine the importance of this assumption, I performed sensitivity analyses using weights of .2/.8. This did affect the results, particularly for one customer. I will discuss this in Section 5.5 under "The Value of Better Forecasts".

Conclusion

Any rate other than full spot prices has errors from at least two sources: aggregating different hours into a single period, and mis-forecasting future spot prices. The first of these can be estimated rigorously from actual historical data, but the second requires a full model of the timing and accuracy of information coming to the forecaster. For example, weather is an important influence on demand; therefore the accuracy of long range weather forecasts will affect the accuracy of system lambda forecasts. Although such analysis may be quite important (see Section 5.5), it is not consistent with the scope or orientation of this thesis or with the

Figure 5.1.5
 Peak and off-Peak Averages of System Lambda
 1979-80

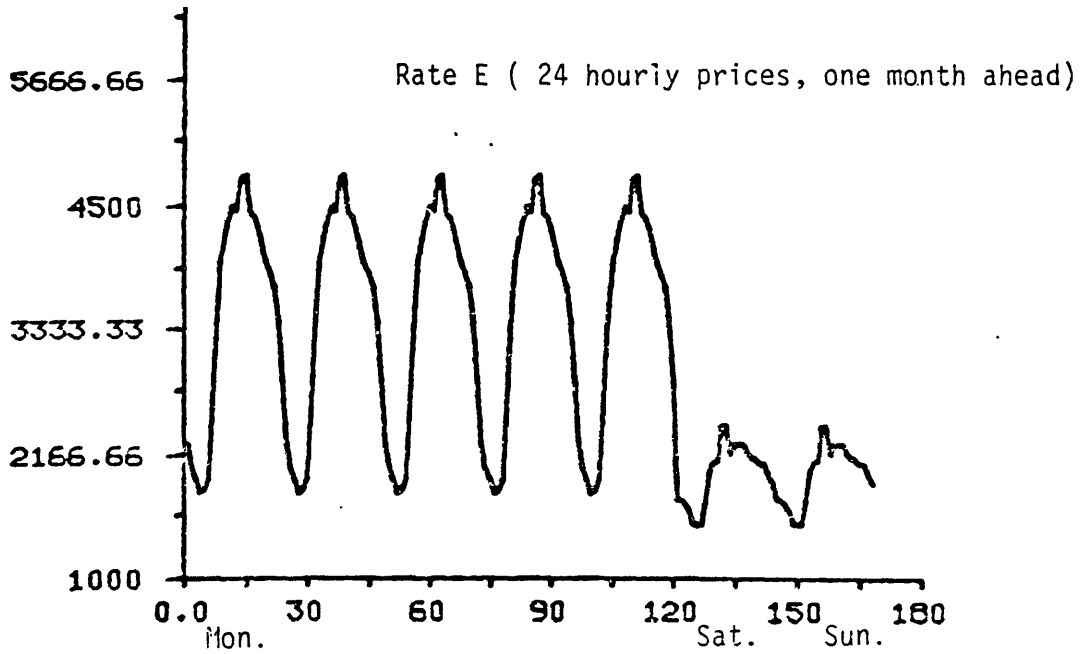
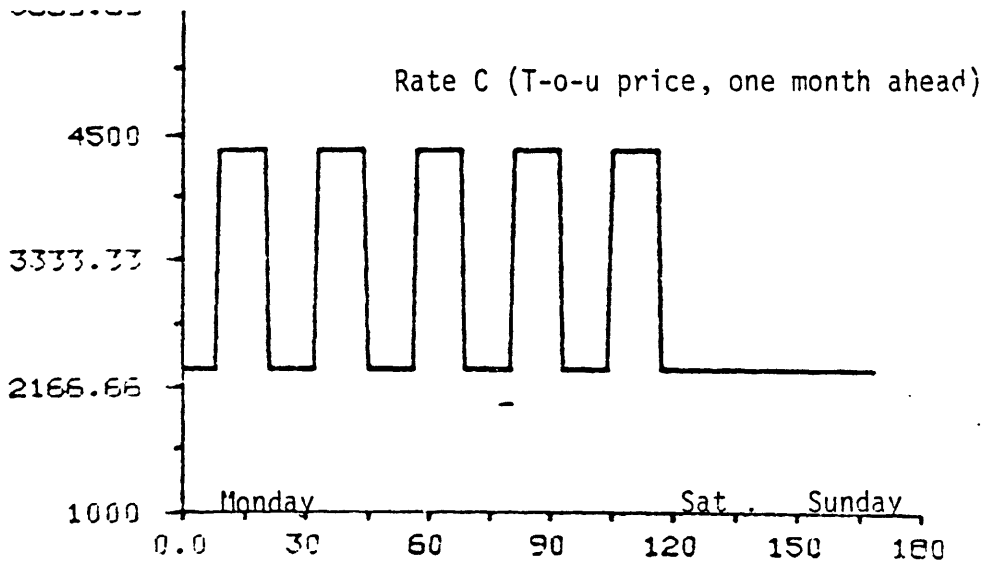


data available on the subject utility. Therefore I have used reasonable but rudimentary simulated forecasting for all of the non-spot rates.

Figure 5.1.6 shows the prices under rates D and E for August. (For comparison, Figure 5.1.1 showed actual full spot prices for the month.) Rate D, like rate B, is a two level time-of-use rate. Rate E has a vector of 24 prices for weekdays, and a different vector for weekends. Rates C and A, not shown, have flat prices for the entire month. Prices under rates C, D and E are recalculated at the beginning of each month. Prices under rates A and B are set only once, for the entire year.

Figure 5.1.6

Prices in August under Two Rates
(in ¢/kwh x 1000)



The prices resulting from each rate are summarized in Figure 5.1.7, which shows the mean and standard deviation of prices under each rate, and the standard deviation of the difference from full spot prices at the corresponding hour.^{1.10}

Figure 5.1.7

Summary of 1980 Prices under Each Rate (£/kWh)

<u>Rate</u>	<u>Mean</u>	<u>Std. Deviation</u>	<u>Std. deviation from full spot</u>
A (Annual flat)	2.499	0	1.35
B (Annual t-o-u)	2.498	0.66	1.16
C (Monthly flat)	2.629	0.35	1.32
D (Monthly t-o-u)	2.629	0.79	1.12
E (Monthly 24)	2.628	0.92	1.02
F (Full spot)	2.631	1.35	0

Despite the large and comparatively consistent difference between full spot prices and the others, it will turn out that different rates have very different effects on social welfare.

5.2 Case Study Customers

For the case studies, four generic customers were modeled and analyzed, each corresponding to different common patterns of electricity using operations. Two of these customers are of the "pure storage" type, as defined in Chapter 4. That is, they use the same amount of electricity each day, but use it at different times depending on the relative prices. They can effectively "store" electricity as another form of energy embodied in intermediate or final products. The other two customers are of the "pure shutdown" type; they have no intertemporal substitution possibilities, but instead simply shut down their operations or switch to alternate fuels if electricity prices go too high. The effects of different rates on each customer are quite different, in part because of this distinction between shutdown and storage behavior.

The customers to be modeled are in no sense a valid sample of the subject utility's electricity using firms. Instead selection of the four customers was based on three criteria, one of which is that they be likely to show responsiveness to time varying prices. The second criterion is that they be simple. This makes the results easier to understand, and holds down computer costs. The third criterion is that each be an idealized version of some actual electricity using process, preferably a widespread one.

The specific customers used for case studies are:^{2.1}

- o A "simple storage" customer which can store exactly one hour of electricity, or its equivalent. Related processes include

some building cooling systems, and municipal water pumping systems.

- o A "discrete rescheduling" customer, which must produce a fixed amount of final products each day, but can schedule a "break period" for several hours at some time during the day, and use no electricity during that period. This fits a variety of electricity intensive production processes, as long as they are not currently operating "flat out." Several examples of this were observed during field trips.^{2.2}
- o A "composite shutdown" customer which can substitute electricity for natural gas or oil to supply clean process heat. Since electricity can usually be used more efficiently at the point of consumption, this may be cost effective even when electricity appears to be more expensive on a per Btu basis. Examples include drying processes where product cleanliness is important.
- o A "standby generator" customer which already owns a diesel standby generator for emergency use, and can use it to save money whenever the spot price exceeds the generator's operating cost. Many hospitals have such equipment sized to provide a substantial fraction of their average load.

I will now give detailed descriptions of the individual customers.

Simple Storage Customer

The first customer is a "pure storage" process as modeled in Chapter 4. Such customers always have a fixed total energy demand over a sufficiently long interval, no matter how high the price. They store some intermediate "product" such as thermal energy, so that busbar electricity demand can be reduced to zero at times of highest price. For simplicity I model a very simple storage process:

- o One hour of storage capacity for the intermediate product.
($s_1 = 1.0$ hours, in the notation of Chapter 4.)
- o Storage can be fully recharged in one hour or less. ($x_1 = 1$ hr.)
- o Exogenous final energy demand corresponding to an average load of 5 MW. ($X_{\min,2} = X_{\max,2} = 5\text{MW}$)
- o No losses in storage. ($L = 1$)
- o One day at a time scheduling. Each midnight the customer sets a production plan for the next 24 hours, without regard to prices 25 or more hours away.^{2.3}

This description is a reasonable approximation of many building heating and cooling systems which use water as a heat exchange medium, and of many thermal processes where temperatures must be maintained within some band.^{2.4,2.5} It also fits a run-of-river hydro generator with a small pond which can store one hour of water flow before release.^{2.6} Municipal water systems usually have much more than an hour of storage capacity, although they are more complex than the simple storage customer because of demand fluctuations.

Optimal behavior for such customers is to look for "peaks" and "troughs" in the upcoming day's prices, and to fill storage during each trough and empty it during each peak. This was diagrammed in Chapter 4. The private value of each megawatt hour of storage is thus at least the difference between the lowest price each morning and the highest price later that day.

Discrete Rescheduling Customer

The second customer is a modified version of a pure storage process. This customer has no binding storage capacity constraint, but has a fixed amount which must be produced each day and only a slight surplus production capacity.^{2.7} By producing "flat out" for most of the day and shutting down for a block of hours at the time of highest price, it can produce at lowest cost. I further assume that all non-producing time must be taken in one block, to minimize internal disruption.

An actual example from a site visit will clarify this.^{2.8} An electric arc furnace is currently open ten hours per day, from 6 AM to 4 PM, and takes 2.5 hours to produce one batch of steel. Final production is thus four batches a day. By operating flat out, the furnace can actually process a batch in 2.0 hours, and shut down for 2 hours at the time of highest price.^{2.9} The case study customer can also take a two hour break, sometime between 6AM and 4PM. It has an electricity demand of 5 MW when operating. Another example is a scrap

metal company with a multi-megawatt crushing machine which is used only about four hours a day.

These firms could also carry intermediate products over from one day to the next, and reschedule production within a week instead of producing the same amount each weekday. This could decrease electricity costs significantly, but to model this would require some probability structure for several day ahead price forecasts under spot pricing. By ignoring this behavioral option, I understate the benefits of spot pricing.

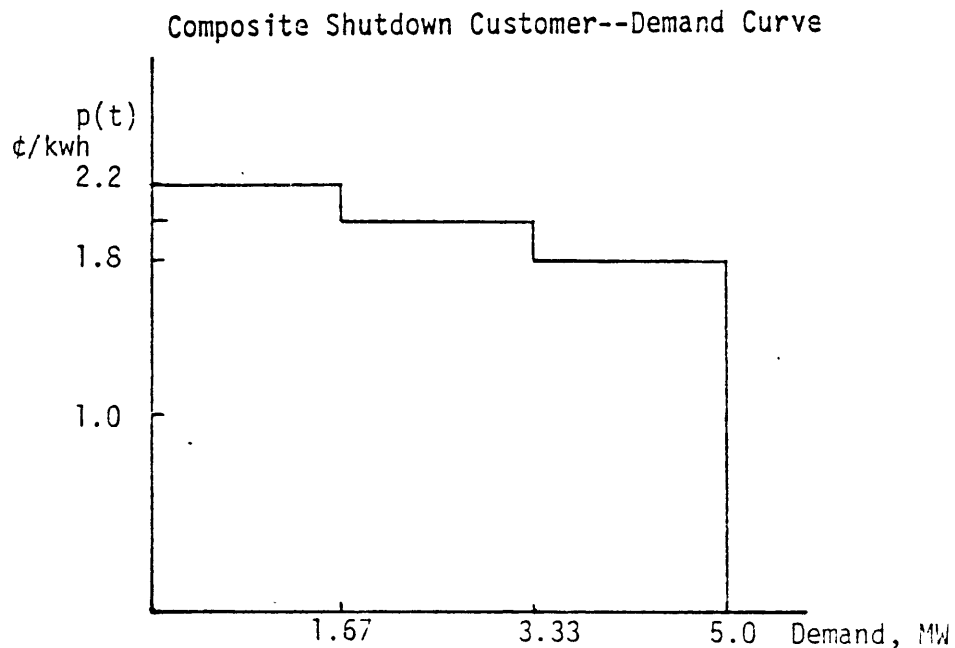
Composite Shutdown Customer

The third customer modeled is a collection of pure shutdown processes. Recall from Chapter 4 that a pure shutdown process simply stops using electricity whenever the electricity price exceeds a threshold. Different uses of electricity lead to different thresholds. For this particular customer I have selected the threshold to correspond to a process which currently provides heat from oil or natural gas for drying, space heating, or processing. If the price of electricity is low enough, the cost of electricity per Btu of delivered thermal energy may be lower than that of the fossil fuel. The exact switchover price depends on the delivered cost of the fossil fuel and on the differential efficiency of electricity versus fossil fuel for the application. For example if fossil fuel costs \$4 per million Btu and has an end use efficiency which is 70 percent of

that of electricity, then the equivalent price for electricity is 1.95 cents per kilowatt hour.^{2.10}

My model assumes no lags or constraints in switching to or from electricity. Thus each hour, the customer compares the current price of electricity with the efficiency adjusted cost of its fossil fuel, and uses whichever is cheaper. I assume the customer has a total load of 5 MW, split evenly among three shutdown prices: 1.8 ¢/kwh, 2.0¢/kwh, and 2.2 ¢/kwh. Thus the customer's instantaneous demand for electricity is shown in Figure 5.2.1. This curve is assumed to hold 24 hours per day, 8784 hours per year.

Figure 5.2.1



Standby Generator Customer

The last case study customer is a customer which, perhaps because of a critical need for uninterruptable power, has installed its own diesel powered standby generation system. Hospitals and airports are obvious examples. A few industrial customers also have such equipment because their manufacturing equipment can be damaged by unexpected outages.

Standby diesel generator systems are comparatively inefficient and use expensive diesel fuel. For example, a heat rate of 11,400 Btu/kwh, a fuel price of \$.88 per gallon, and maintenance costs of 0.9 cents per kwh lead to a cost of 8 cents per kwh generated.^{2.11} I therefore use a price of 8 cents per kilowatt hour as the shutdown point, i.e. the point at which the diesel is put on line. I model a generator of 5 MW output.

The spot prices discussed in Section 5.1 exceed 8 cents per kilowatt hour for only 51 hours during the year, and have a maximum of 11.5 cents per kwh. Therefore the standby generator customer will "shut down" (i.e. use its generator) only a negligible amount under these prices. I will therefore not report results on this customer until the higher price scenario, in Section 5.5.

5.3 Results

Given the rates described in Section 5.1 and the customers described in Section 5.2, we can use the models of Chapter 4 to predict the hour by hour behavior of each customer under each of the six rates. We can then compare behavior under the different rates to find their relative effects on the customer's own profit, as well as on social welfare. Specific quantities of interest include:

- o The customer's own costs for electricity plus other factors of production, under each rate.
- o The social value of the electricity and other factors of production used by the customer. The difference between private costs and social value is the cross-subsidy received by the customer under the rate.
- o Which rate is socially optimal for the customer when transactions costs are considered.
- o Which rate the customer would select if allowed to choose.
- o What incentives the customer would have under each rate to make various investments which change its electricity use.
- o The social value of these investments under different rates.

I will discuss the results for one customer in detail, to show how the analysis is done. I will then summarize the results for the other customers, and discuss the important implications. Incentives for investment under the rates will be covered at the end of the section.

Assigning customers to rates will be mentioned here, but thorough discussion will be in Section 5.4, where transactions costs are estimated.

"Simple Storage Customer" Results

The simple storage customer uses 120 megawatt hours of electricity every day, or 43,920 MWh a year. By taking advantage of its one hour of storage capacity it can purchase more of this electricity during the hour of lowest price each day, and none during the hour of highest price. Note that under flat prices (rates A and C) it will have no incentive to do so, hence its storage capacity is irrelevant under these rates.

Figure 5.3.1 shows the behavior of the customer under each rate.

Figure 5.3.1

Results for Simple Storage Customer--Detail

Rate	Hours Purchasing Electricity	Hours Using Storage	Cost of Electricity Used	
			Cost to Society*	Cost to Customer*
A (Annual flat)	8784	0	1155	1098
B (Annual t-o-u)	8528	256	1152	1080
C (Monthly flat)	8784	0	1155	1154
D (Monthly t-o-u)	8528	256	1152	1135
E (Monthly 24)	7614	1170	1121	1119
F (Full spot)	7847	937	1114	1114

*Costs in thousands of dollars per year.

The first column shows how many hours the customer purchased electricity; the second column, how many hours it "coasted" on its storage. For example under two-level time of use rates (B and D) storage was used exactly one hour each (non-holiday) weekday, a total of 256 hours.

The third and fourth columns show the cost of the electricity used. The third column values the electricity at its social resource value, which is, by definition, the full spot price. This social cost is of course minimized under rate F, i.e. when the customer is on full spot prices. Full spot prices lead to a gross social savings from this customer's behavior of 41 thousand dollars per year, compared with its behavior under flat prices ($1155-1114=41$). It is interesting to note that rate E gives 80 percent as much savings as rate F, for this customer. Nonetheless it is probably socially preferable to put this customer on full spot prices. (This will still be true when we consider transactions costs in the next section.)

The fourth column takes the customer's perspective, and values the electricity at its price to the customer under each rate. Thus under rate B the customer would have paid only \$1,080,000 for electricity which had a social value of \$1,152,000. Thus the customer was subsidized by \$72,000 under rate B. Some of this difference stems from the fact that when prices under rate A were set, spot prices for 1980 were forecast too low by 5 percent. This accounts for $1080 \times (.05/.95) = 57$ thousand dollars of the subsidy, and this amount could not have been anticipated by the customer. Thus the anticipated subsidy to the customer under rate B was only $72-57 = 15$ thousand

dollars. (The exact amount from more precise calculations is closer to \$16,000.)

We can collect this information to compare different rates, as in Figure 5.3.2.

Figure 5.3.2
Social and Private Savings Due to Time Varying Prices
(Simple Storage Customer)

<u>Rate</u>	<u>Social Savings Over Rate A</u>	<u>+ Antici- pated Subsidy</u>	<u>= Gross Anticipated Private Savings Over Rate A</u>	<u>- Approx. Trans. Costs</u>	<u>= Net Antici- pated Private Savings</u>
A (Annual flat)	0	0	0	0	0
B (Annual t-o-u)	3	16	19	0	19
C (Monthly flat)	0	1	1	0	1
D (Monthly t-o-u)	3	16	19	1	18
E (Monthly 24)	35	2	37	2	35
F (Full spot)	41	0	41	4	37*

* = Best Rate.

The first and third columns give the gross social and private anticipated savings for this customer from being under each rate, compared with being on rate A. We can see that rate F is the best rate both for society and for the customer's own profits, before considering transactions costs. The fourth column shows approximate transactions costs (from Section 5.4). The fifth column shows that if these are correct transactions costs for this customer, the customer would find rates E and F almost equally profitable, and prefer one of

them to any of the others. Although it is not shown, an analogous calculation would show that society is best off if the customer is on Rate F. (Net social value = Column 1 minus Column 4).

Figure 5.3.3 shows the results for the simple storage customer in a slightly different format. Its first and third columns show the social and private costs of all energy related resources used for the year.^{3.1} These are taken from the last two columns of Figure 5.3.1. The second and fourth columns of Figure 5.3.3 correspond to the first two columns of Figure 5.3.2, and show the relative social value and private subsidy of each rate. The final column is the expected relative private value of each rate, from the third column of Figure 5.3.2. Customers would select the rate with the highest private value, minus any adjustment for transactions costs.

Results for Other Customers

Figures 5.3.4 and 5.3.5 present the results for the discrete rescheduling customer and the composite shutdown customer respectively. Each row corresponds to the effects of the corresponding rate.

Several patterns are visible in these results. I will discuss the social welfare implications, then the implications for the customers' private profits.

Figure 5.3.3
 Results for Simple Storage Customer
 (Thousands of Dollars Per Year)

<u>Rate</u>	<u>Total Social Cost</u>	<u>Relative Social Value</u>	<u>Realized Private Cost</u>	<u>Ex ante Expected Subsidy</u>	<u>Expected Private Value</u>
A (Annual flat)	1155	--	1098	0	0
B (Annual t-o-u)	1152	3	1080	16	19
C (Monthly flat)	1155	0	1154	1	1
D (Monthly t-o-u)	1152	3	1135	16	19
E (Monthly 24)	1121	35	1119	2	37
F (Full spot)	1114	41	1114	0	41

Gross social savings of full spot pricing over current rate B: \$38,000 per year.

Privately preferred (ex ante) rate: F

Socially preferred rate: F

Assumptions: Pure storage process, no shutdown

Normal operating rate: 5 MW, 24 hours per day, 8784 hours per year
 Storage equivalent to: 5 MW hours
 Storage discharge time: 1 hour
 Maximum upstream operating rate: 10 MW
 Storage recharge time: 1 hour

Figure 5.3.4

Results for Discrete Rescheduling Customer
(Thousands of Dollars Per Year)

<u>Rate</u>	<u>Total Social Cost</u>	<u>Relative Social Value</u>	<u>Realized Private Cost</u>	<u>Ex ante Expected Subsidy</u>	<u>Expected Private Value</u>
A (Annual flat)	444	--	366	59	59
B (Annual t-o-u)	440	4	401	18	22
C (Monthly flat)	444	0	384	59	59
D (Monthly t-o-u)	440	4	422	18	22
E (Monthly 24)	435	9	434	1	10
F (Full spot)	431	13	431	0	13

Gross social savings due to full spot pricing over current rate B:
\$9,000 per year.

Privately preferred (ex ante) rate: any flat rate

Socially preferred rate: F

Assumptions: Pure storage process, no shutdown.

Normal operating rate: 5 MW, 8 out of 10 hours per day, 2928
hours per year.

Process can shut down for 2 hours each day out of the 10 open.

The hours must be contiguous.

Customer is open 6 AM to 4 PM daily.

No storage of product from one day to the next.

Figure 5.3.5

Results for Composite Shutdown Customer
(Thousands of Dollars Per Year)

<u>Rate</u>	<u>Total Social Cost</u>	<u>Relative Social Value</u>	<u>Realized Private Cost</u>	<u>Ex ante Expected Subsidy</u>	<u>Expected Private Value</u>
A (Annual flat)	878	--	878	0	0
B (Annual t-o-u)	871	7	860	0	7
C (Monthly flat)	878	1	877	1	1
D (Monthly t-o-u)	861	18	855	5	23
E (Monthly 24)	830	49	824	6	54
F (Full spot)	791	87	791	0	87

Gross social savings due to full spot pricing, over current rate B: \$80,000 per year.

Privately preferred (ex ante) rate: F

Socially preferred rate: F

Assumptions: Pure shutdown process

Maximum operating rate: 5 MW, 24 hours per day, 8784 hours per yr

Shutdown points: one third of load shuts down above 1.8 ¢/kwh, two thirds above 2.0 ¢/kwh, all above 2.2 ¢/kwh

Costs include electricity use plus cost of alternate fuel used while "shut down". (If no electricity used, fuel costs 20 x 5 x 8784 = \$878,400 per year.)

Short Term Social Welfare

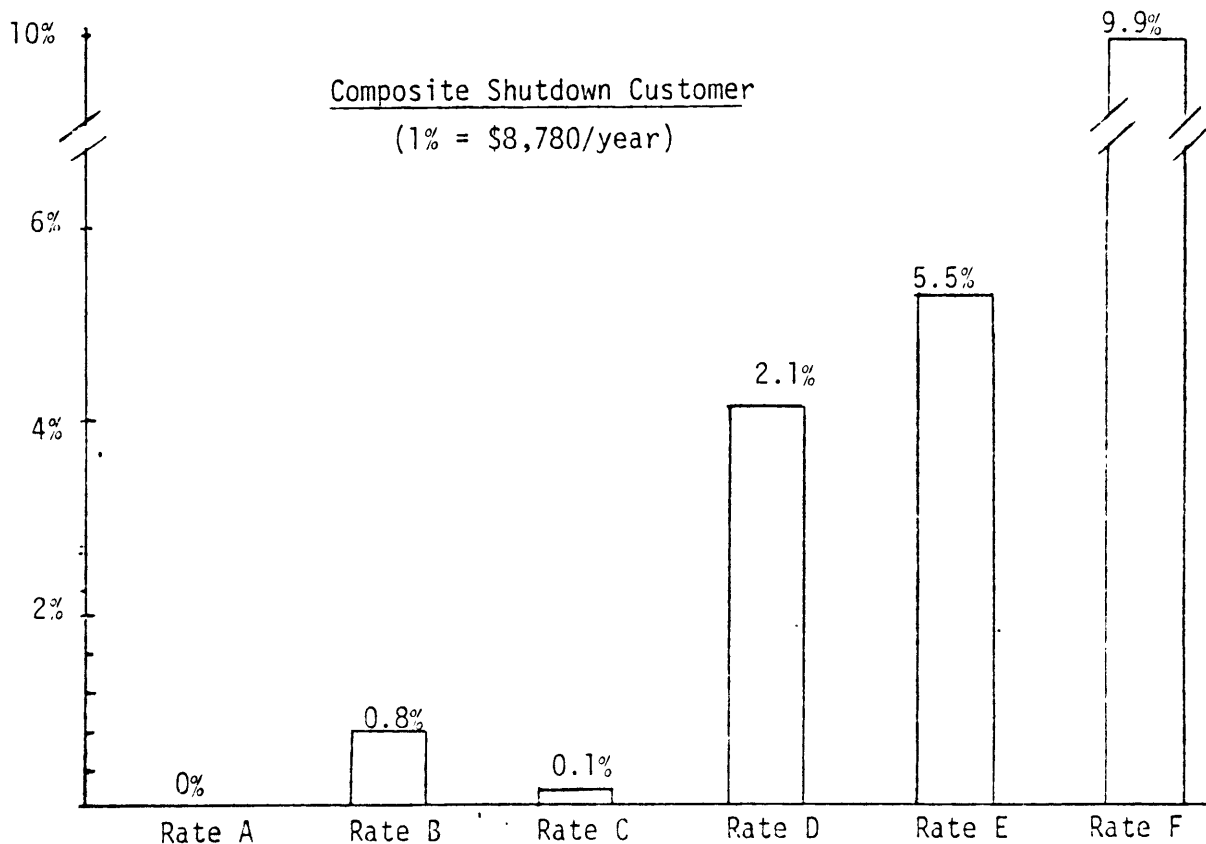
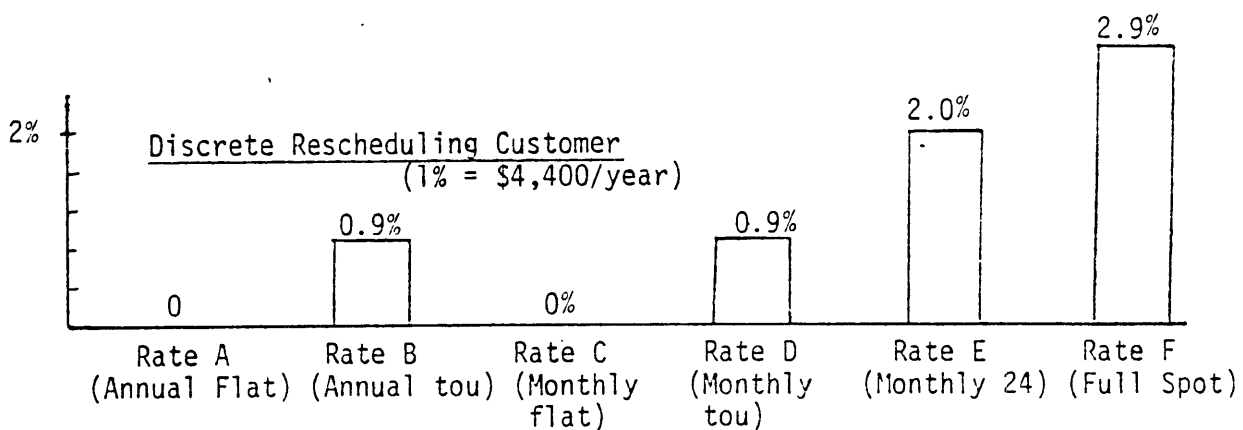
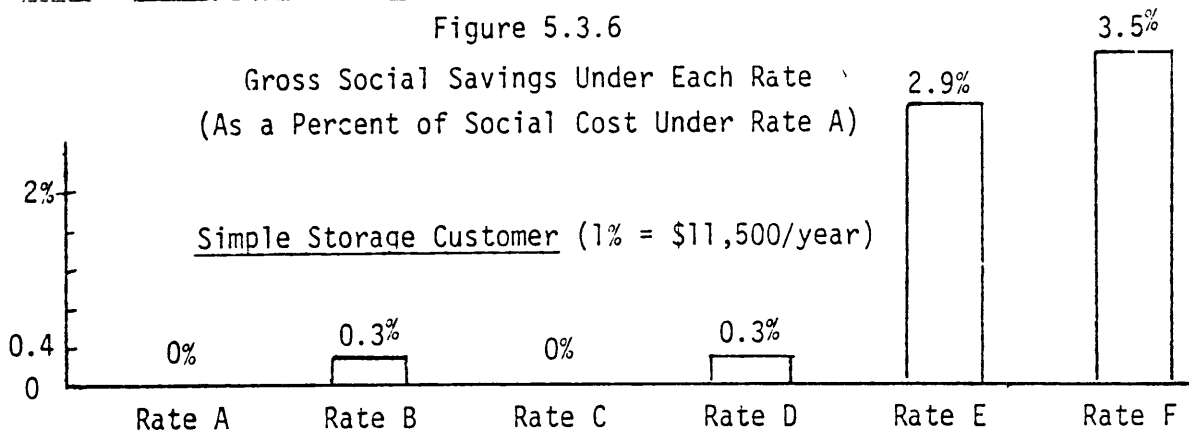
One important consequence of each rate is the social welfare value of customer behavior under that rate. I have modeled each customer as a cost minimizer with fixed final production; therefore social welfare maximization is equivalent to social cost minimization. Energy related costs are almost always highest for rate A (annually adjusted flat prices), since its prices have the greatest deviation from full spot prices. They are lowest for rate F, and intermediate for other rates.

The second column of Figures 5.3.4 to 5.3.6 is the relative social value of each rate, measured as savings beyond the cost incurred under rate A. These measurements are before adjustment for transactions costs. Figure 5.3.6 shows the relative social savings under each rate, as a percent of the social costs under flat prices (rate A). Several results are visible:^{3.2}

- o Annually recalculated two-level time-of-use rates (rate B), which are currently applied to industrial customers in this utility jurisdiction, yield positive but relatively small gross social benefits, of less than one percent of total costs.^{3.3}
- o Full spot pricing (rate F) yields much larger social benefits than does rate B. The savings are very customer specific; for the cases studied here they are three to ten percent of social costs. Of course the savings net of transactions costs will be smaller, especially for small customers, and

Figure 5.3.6

Gross Social Savings Under Each Rate
 (As a Percent of Social Cost Under Rate A)



are discussed later.

- o Monthly recalculation of 24 hourly prices (rate E) yields net social savings which are much larger than two-level time-of-use prices. For storage customers, it gives savings on the order of three quarters of those under full spot pricing. This was unexpected. The reason is that optimal storage behavior requires knowing mainly the time of system price peaks, not their level. Prices under rate E have either the correct time of peak on a given day, or a time with a price close to that day's peak.

Performance of this rate for shutdown customers is not as good since what drives their behavior is extremes in the level of prices. Any rate other than full spot prices will smooth the extremes, reducing benefits.

- o Any two-level time-of-use rate (rates B or D) will have the same social cost effects on pure storage customers as any other such rate, provided that the times at which prices change are the same. As long as the morning price change is large enough to induce the customer to set up a response pattern, any such rate will lead to the response pattern of storing intermediate product just before the change and using it just after.^{3.4}
- o The savings for either of the storage customers would be almost the same if they operated only weekdays, since the trough-to-peak spot price change is highest on weekdays.



Room 14-0551
77 Massachusetts Avenue
Cambridge, MA 02139
Ph: 617.253.5668 Fax: 617.253.1690
Email: docs@mit.edu
<http://libraries.mit.edu/docs>

DISCLAIMER OF QUALITY

Due to the condition of the original material, there are unavoidable flaws in this reproduction. We have made every effort possible to provide you with the best copy available. If you are dissatisfied with this product and find it unusable, please contact Document Services as soon as possible.

Thank you.

*Author mis-type error. Page 296
doesn't exist.*

under these rates, since the average full spot price from 10 AM to 5 PM is higher than from 8 AM to 8 PM. Either rate E or rate F would remove subsidies for this type, since these rates do not have any time aggregation error.^{3.5}

- o Time varying rates will cause customers to alter their pattern of electricity use. This alteration will create a second cause of subsidies, if it saves the customer more money than the change in the value of the electricity it uses. Specifically, under two-level time-of-use rates the storage customers save the difference between the price at 7 AM, and the price at 8 AM. This is usually much less than the difference between the full spot prices for those hours, leading to the subsidies for those customers under rates B and D.

For the customers I have modeled (including some not discussed here), it appears that only the subsidies under flat prices are really extreme. Even two-level time-of-use pricing reduces the subsidies to a few percent of the electric bill.^{3.6} Customers will mis-select a rate only when its subsidies outweigh its higher social costs. (See the last column of Figures 5.3.4 to 5.3.6. The lowest cost gives the rate the customer will select.) Thus if customers are given a choice between full spot prices (or rate E) and two-level time-of-use prices, they will select time-of-use prices only when the social savings of full spot prices are small. Only if flat prices are an option is mis-assignment a large problem.

As discussed in Section 5.1, under some rates the cross-subsidies for all of a utility's customers on the rate may not sum to zero. This is particularly likely for rates A and C. In such cases the prices under the rate will, in practise, have to be adjusted. How will this affect the results?

Suppose the adjustment is done by increasing all prices under the rate by a constant. Then the behavior of the storage customers will be the same as before since only relative prices affect their behavior. Shutdown customers will be affected if the change pushes any price over its shutdown threshold. If so, the social cost of the customer's behavior will unambiguously be increased, since the discrepancy between prices under the rate and full spot prices will increase.

Hence such a subsidy balancing price adjustment will make full spot prices socially more attractive. Whether spot prices will become more attractive to the customer depends only on whether the prices under the other rate were increased or decreased.

The Value of Capital Investments

Which rate a customer is under affects both the social and private value of capital investments. These values are defined as the change in short term costs as a function of the new investment; see Chapter 3 for a general discussion of optimal investment. Full spot pricing provides socially optimal investment incentives in two regards.

First, the social and private values of an investment must be equal under full spot pricing, but may not be equal under other rates. Second, the social value of a given investment is usually higher under full spot prices than under any other rate.^{3.7} It will also usually be true that the customer's own incentives to make an investment will be higher under full spot prices.

The case study customers have a variety of investment possibilities, each of which may have different value. I will discuss several generic types of investments, rather than each possibility.

The first type of investment is a linear scale up of an entire customer's plant, which increases both electricity using equipment and final production capacity by the same amount.^{3.8} Then the customer's short term cost function will increase proportionally. To the extent that electricity costs for the original plant size are different under different rates, so will the net value of additional output produced by the plant expansion be different. For example, for the composite shutdown customer, Figure 5.3.5 shows that doubling plant size and output will increase social costs by \$38,000 per year less if the customer is under full spot prices than it would under rate E. In short, plant expanding capital investments are at least as socially valuable under full spot prices as under any other rate. Unless the social value of full spot prices is outweighed by a subsidy, which it is not for this customer, these investments are also more privately profitable under full spot prices.^{3.9}

The second type of investment is one which holds constant the customer's final production capacity, but adds the option to use electricity differently over time. An example is constructing a run-of-river hydro site with or without a pond which can be used to hold back an hour of water. Figure 5.3.7 shows the social and private values of the investment in storage equipment for the simple storage customer. The social and private values of the investment are higher under full spot prices than under any other rate.

Figure 5.3.7

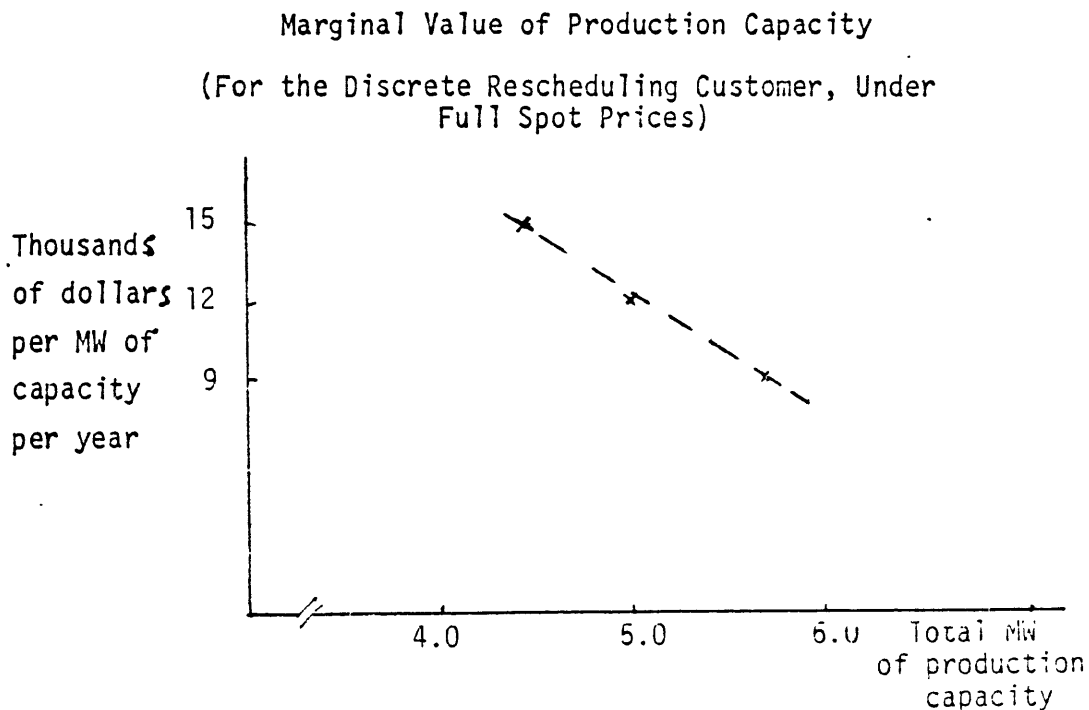
Total Value of an Investment in Storage Capacity
(Thousands of Dollars per Year)

<u>Rate</u>	<u>Social Savings</u>	<u>Private Savings</u>
A (Annual flat)	0	0
B (Annual t-o-u)	3.1	18
C (Monthly flat)	0	0
D (Monthly t-o-u)	3.1	18.5
E (Monthly 24)	35	35
F (Full spot)	41	41

Explanation: Each number is the difference between the cost of electricity with one hour of storage capacity (see Figure 5.3.1 for details) and the cost of electricity without the storage capacity, assuming that once installed the storage capacity is used to maximize the customer's profits. The first column values electricity at its social value; the second is the profit improvement to the customer, under the prices actually charged.

Finally, what is the incremental value of an expansion to one portion of an electricity using system? Additional storage or recharging capacity becomes less valuable, the more is in use. Figure 5.3.8 shows this effect for an investment in upstream production capacity by the discrete rescheduling customer.^{3.10} The larger its upstream capacity (the MW rating of its melters, for the arc furnace) the fewer hours it needs to produce the same daily output, and therefore the more flexibility it has to schedule work at the time of lowest electricity prices. Increasing capacity from 4 MW to 4.44 MW decreases work time by 1 hour, and under full spot prices saves \$15,000 per year per MW of capacity. Further investments yield diminishing marginal value, as shown. This general behavior is true for any component of a multi component plant. It holds no matter what rate the customer is on.

Figure 5.3.8



5.4 Transactions Costs, Customer Size, and Assignment to Rates

A key concept is that the improvement in conventional welfare (what I have called gross social welfare) resulting from more sophisticated rates must be compared with the increased transactions costs of such rates. (Section 3.5 formalized this relationship.) Customers which have large absolute (kwh) responses to time varying prices should be put on full spot prices or rates close to it. Customers can have a large absolute response to prices either because they have large total loads, or because they shift a large percentage of their load in response to price signals. Section 5.3 measured the gross value of responses to different rates for 5 MW customers of four types. In this section I will compare these with transactions costs, and determine the social welfare under rates A through F.

The end result of the calculations in this section is that, depending on the customer type, the "breakeven size" for full spot pricing is between 0.5 and 5 MW; even larger in a few cases. Customers above the breakeven size should be put on full spot pricing.

Estimating Transactions Costs

The first step is to estimate the per customer transactions costs associated with the rates. These must be estimated as a range, since they will be site specific. Also, some of the technology involved is new and somewhat speculative, making cost estimates less reliable.

The transactions costs of each rate are of five types:

1. Metering costs
2. Billing costs
3. Informing customers of the prices
4. Customer costs of responding to the prices
5. Potential opportunism by the utility.

I will assume that potential opportunism is the same for all rates, for one of the reasons discussed in Section 3.6. Notice that the third and fourth costs are actually discretionary: if a customer decides their value is less than their cost, it may elect not to incur them.^{4.1}

For example, under rate E (monthly recalculation; 24 hourly prices for weekdays and another 24 for weekends), incremental transactions costs per customer are the costs of:

1. Metering the customer's demand. A recording demand meter is at present the most practical way to do this, although theoretically a simpler meter could be used.
2. Calculating and confirming the customer's bill each month.
3. Mailing or otherwise communicating the next month's prices, once a month.
4. The customer's computational costs for deciding how to behave, and internal costs of implementing that decision. These costs will be of two types: an initial cost of devising the algorithm to determine behavior, and a monthly cost of executing that algorithm.^{4.2}

Information on these costs is sketchy. The best available is for the many residential time-of-use pricing experiments and

implementation. Relevant information from these and other sources is covered below.

Billing (including meter reading):

- o Monthly billing costs under flat rates (A or C) would be the same as they are today, since bills are already calculated using fuel adjustment charges which change regularly. Define this cost as the point of comparison.
- o Sturgeson [1980] reports on several time-of-use pricing experiments. Monthly billing under two-level time-of-use rates (B or D) is initially more expensive than conventional billing, because of higher error rates, customer complaints, and other problems. Average costs during the shakedown period were estimated as approximately \$10 and \$13 per month in two states, and "twice those of ordinary billing" in another [Sturgeson, 1980, vol. 2, p. 81]. Incremental cost per customer after the shakedown period would presumably be lower. Note that these costs were for magnetic tape based metering, whereas for rates B and D two dial time of use meters would be adequate. Therefore \$3 to \$15 per customer per month probably covers the correct additional costs.
- o Monthly billing costs under rates E or F would be higher since essentially hour-by-hour calculations are required, and missing data (a common problem with magnetic tape recording meters) requires more human judgement to estimate a bill. The above numerical estimates in Sturgeson should therefore be raised. Doubling them gives about \$25 per customer per month during the shakedown period. \$10 to \$25 per customer per month gives a reasonable range.

Metering Costs:

- o Rates A and C can use conventional meters.
- o Two-dial time-of-use meters (with battery backup) are estimated as \$140 per point without installation in Sturgeson [1980, vol. 1, p. 11] and \$232 per point with installation in Ebasco Services [1977]. A range of \$200 to \$300 per point for capital costs is reasonable.
- o Magnetic tape recording meter costs are given as \$300 to \$600 per point in 1975 in Sturgeson [1980, vol. 1, p. 15] \$600 per point in Ebasco Services [1977], and \$800 per point in Gorzelnik [1979]. Since these meters are very unreliable, double meters are needed for industrial customers. Inflating to 1980 costs then gives about \$100 to \$2000 per point.

Communicating Prices:

- o Costs under annual rates (A and B) will be taken as the point of comparison.
- o Under monthly rates (C,D, and E) a monthly mailing (or publication in newspapers) are needed. Since the mailing need not be personalized a cost of \$1 to \$5 per customer per month is reasonable.
- o Under full spot pricing, hourly communication is needed. I visited one customer in San Diego which built a device to decode digital telephone signals, for about \$2000. Today a microcomputer with display screen, modem, and a port for interfacing to hardware costs \$1500 or less. Some link between the customer and the utility is also needed. A dedicated telephone line is always available, but its costs are extremely site dependent. A dedicated telephone line is not necessary, however. A microcomputer can be programmed to dial up a utility recording each hour and "read" the current price. Such a system requires a dedicated telephone set (\$20 to \$40 per month), charges for using the telephone (which depend on distance to the utility) and at most \$500 for additional computer hardware and software. Thus costs could range from \$1500 to \$2500 initial cost and \$20 to \$200 per month for telephone charges.^{4.3}

Customer Decision Making:

These costs will depend entirely on what loads the customer controls in response to prices, and how those loads are currently turned on and off (manually, by self contained controllers of some kind, or by an integrated process control system). For certain kinds of loads such as pure shutdown loads, optimal demand patterns can be calculated easily and immediately once prices are known, even under full spot pricing. For others, optimal response will depend on variables such as outside temperature or order backlog, requiring more expensive decision making.

I will describe the costs of responding to full spot prices, for a hypothetical customer like the simple storage customer described in Section 5.2. Such a customer needs to know only whether the current price is higher or lower than next hour's price.^{4.4} This immediately translates into a decision to draw down storage if next hour's price will be lower, or charge it up if next hour's price will be higher.^{4.5} If the load is already controlled by computer, for example for a building air conditioner, this algorithm can be programmed and debugged in one man day, for a cost of roughly \$500. If the load is manually controlled, the algorithm must be executed

manually each hour, which will require at most one man hour per day at \$20 per man hour, of \$600 per month.

Under rate E (prices which change every hour but are preset a month in advance), the costs of computerized control are essentially the same. But manual control can now be based on a simple schedule ("Turn the load on at these times, and off at these other times"). Such a schedule can be set the first day of the month, and implemented in perhaps one fifth the time each day required under full spot pricing.^{4.6} This implies a monthly cost of \$120 per month, if manual control is used.

Under the two time-of-use rates, the necessary control actions could be taken in one tenth the time required under full spot pricing, since prices change twice a day. Under the flat rates (A and C) no daily actions would be needed, but under rate C action would still be needed each month.

This example has several implications for this component of transactions costs. First, manual decision making can lead to significant costs under full spot pricing, moderate costs under rate E, and low costs under the other rates. Second, these costs will be lower for customers which already have computerized real time information processing or control.^{4.7} But recall that all of these costs are discretionary. That is, a customer under full spot pricing has the option of acting as if it is on rate E, thus reducing its transactions costs. It will act this way if the savings from more sophisticated control are not large enough.

Figure 5.4.1 summarizes these transactions costs. These estimates are very tentative, and can easily be off by a factor of two in specific cases. The calculations use a 30 percent capital charge rate for fixed costs, which reflects the less than ten year average lifetime of most of the capital items.

Figure 5.4.1

Summary of Estimated Transactions Costs
(Dollars per year)

<u>Rate</u>	<u>Non-Discretionary Costs</u>	<u>Total Costs</u>	<u>Sample Point Estimate</u>
A (Annual flat) *	0 *	0 *	0*
B (Annual t-o-u)	95-280	350-1600	400
C (Monthly flat)	5-60	130-300	150
D (Monthly flat)	100-340	550-1750	800
4 E (Monthly 24) 4	420-900	1230-3600	2000
F (Full spot)	420-900	2400-16500	4000

* Costs under rate A are used as the base point for comparisons.

Note: See text for derivations. Costs will be case specific and may fall outside the bounds indicated here. Costs at the low ends correspond to customers with computerized process control already in use. Costs at the high ends correspond to manual control systems. Sample point estimates in the third column are for illustrative purposes and are used in subsequent calculations.

Caveat: These costs are for present, off-the-shelf hardware. Mass-produced systems designed for spot pricing would reduce them, especially the non-discretionary component of costs.

The column labeled "non-discretionary costs" gives the annual cost of all items which the utility must pay for when a customer is assigned to that rate, namely metering, billing, and for rates A through E, communicating prices. Total costs are the non-discretionary costs, plus costs the customer may choose not to incur, if it is willing to sacrifice the additional private profits of being on that rate. Under spot pricing, these are the costs of receiving real time price signals, and of responding to them.^{4.8}

From Figure 5.4.1 we see that the costs of responding to spot prices can be very high for some customers. This reflects the fact that some processes may be too complex and decentralized to make real time control feasible, even if they would be good candidates for spot pricing in a zero transactions cost world.

For specific calculations of net benefits I will use the figures in the last column of Figure 5.4.1. These are in no sense expected values of actual transactions costs for a particular customer. Rather, they are reasonable point estimates for a customer which already has somewhat automated control of electricity use.

Customer Size and the Net Value of Spot Pricing

Using these estimates of transactions costs, we can calculate the net value of each rate for the case study customers. For a given type of customer, transactions costs are essentially independent of its size, while gross benefits of different rates are proportional to size. Hence even if full spot prices are socially optimal for three of the four case

study customers, they may not be for smaller versions of the same customers.

Figure 5.4.2 shows the social benefit of each rate for "simple storage" type customers of different sizes. Benefits are measured relative to flat prices (rate A), as in the second column of Figure 5.3.3. They are shown net of transactions costs, using the last column of Figure 5.4.1 as a point estimate of transactions costs. For example full spot prices are socially inferior to flat prices for customers of this type which are smaller than 0.5 MW. Full spot prices are socially preferred over all other rates only for customers larger than 1.7 MW.

Because some of the transactions costs are discretionary, such "breakeven" sizes should be interpreted with caution. Assigning customers larger than the breakeven size to a less sophisticated rate will cause a loss of net social surplus. But the reverse is not necessarily true. Assigning a 1 MW customer to full spot pricing will not cause a loss of surplus, since the customer would just act as if it were on rate E instead. Mandatory full spot pricing is not actually socially harmful for customers larger than about 250 kw, the breakeven point between rates B and E.

Figure 5.4.3 shows the socially optimal rates for each customer size for three customer types modeled. The standby generator is not shown. From Section 5.3 we know that full spot pricing greatly dominates even monthly adjusted 24 hour pricing (rate E) for 5 MW composite shutdown customers. Full spot pricing has a correspondingly large zone of optimality for this type of customer: customers from 250 kw upward

Figure 5.4.2
Net Social Savings by Customer Size and Rate Used

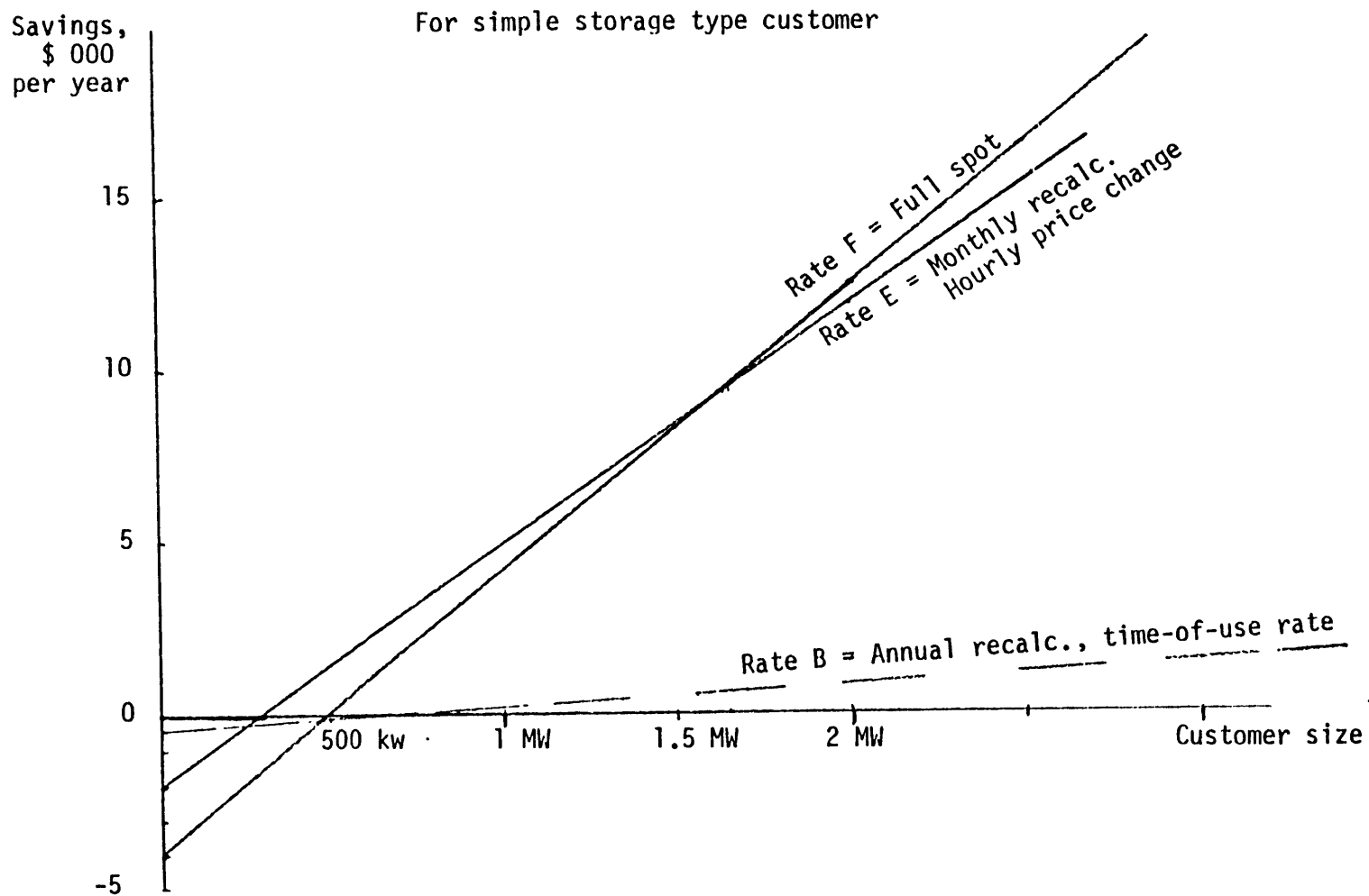
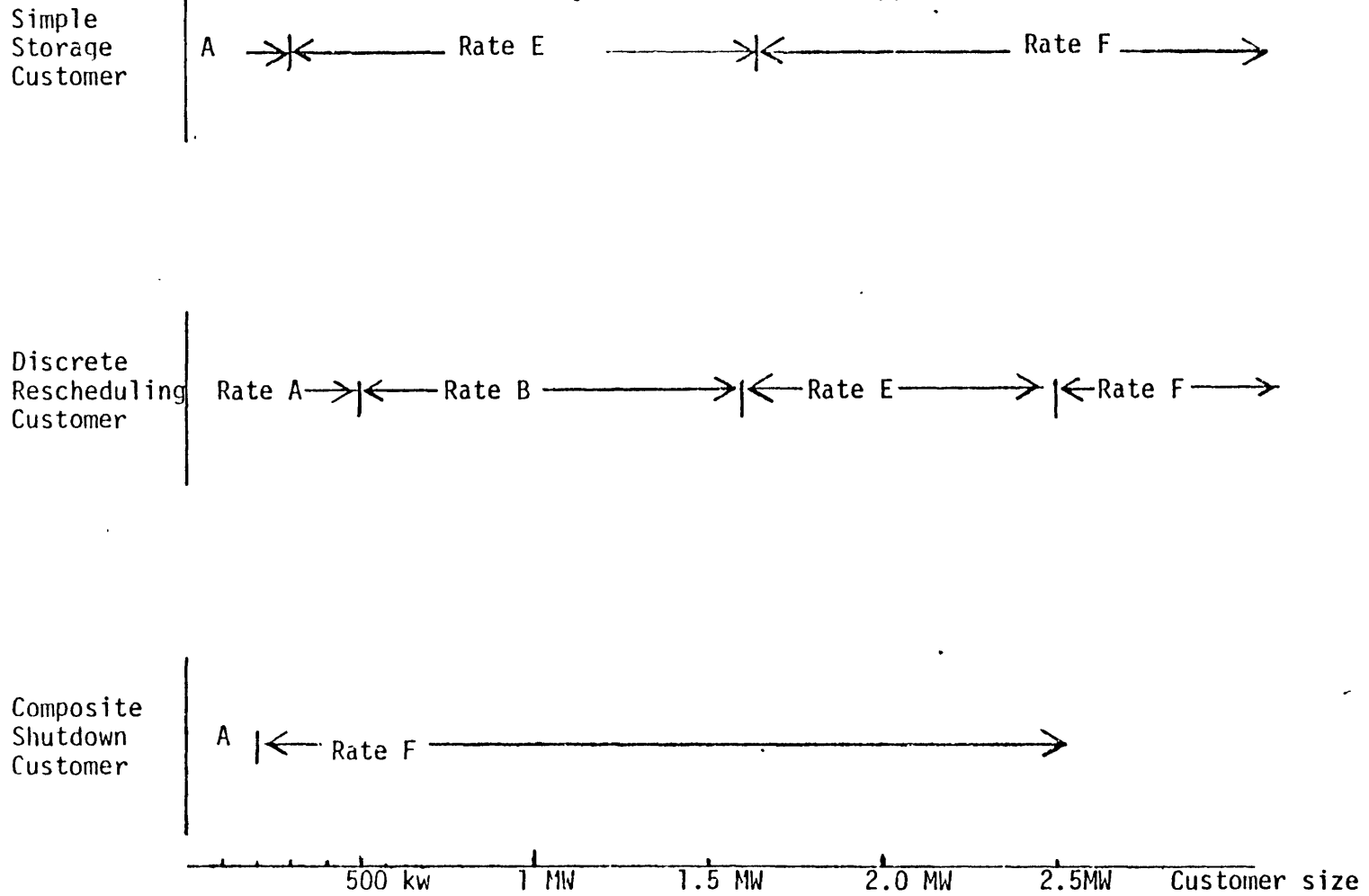


Figure 5.4.3

Socially Optimal Assignment to Rates
by Customer Size and Type



should be on full spot pricing. For other customer types, full spot prices are not optimal below several MW. They are not actually harmful for any of these customers larger than 1.5 MW. Naturally this figure will increase for customers with higher transactions costs.

In Section 5.6 I will show that the subject utility sells about 22 percent of its energy to customers larger than 2 megawatts peak demand.

5.5 Changes in Utility Characteristics

The Midwestern utility described in Section 5.1 is not representative of all utilities in the U.S. This section explores the effects of utility characteristics on the benefits of different time varying rates. The benefits of spot pricing are driven by the variability of the underlying full spot prices. In this section I model a utility very similar to the base case utility, but with more variable full spot prices. The effect is to increase the social value of putting most customers on full spot prices. It also has some effect on the social values of other time varying rates. It has relatively little effect on the cross-subsidies received by customers under various rates. I analyze one kind of change in full spot prices; other changes will have different effects, as I discuss at the end.

I will start by presenting the change made to the base case utility, then present the effects on the different customers. Then I will show the value of better forecasts under some of the non-spot rates. Finally I will qualitatively discuss what kind of utilities will find full spot prices most valuable.

The "With Curtailment Premium" Case

As discussed in Section 5.1, all previous calculations have been based on the rudimentary spot price formula $p^*(t) = \lambda(t)$. Hence the spot prices ignored the effect of marginal losses and curtailment premia on optimal full spot prices. For the sensitivity analysis I simulate the effect of adding a curtailment premium $\mu(t)$ to the full

spot price. This raises prices under all six rates, since the non spot rates are expected values of full spot prices.

To simulate the effect of the curtailment premium, I used the base case system lambda but increased all prices above 6 cents per kwh. Prices this high are likely to involve "emergency" transactions with neighboring utilities, or purchases over very long distances. Under full spot pricing such a curtailment premium might be appropriate instead, especially if the utility were not intertied with neighbors. Such a premium gives the utility an alternative to brownouts or blackouts.

I set the premium to $2(p^*(t)-6)^2$. For example, an original price of 9 cents per kwh goes up to $9 + 18 = 27$.

Figure 5.5.1 shows the effect of this change on full spot prices of the original and the fictitious utilities. The highest price for the new utility is about 72 cents per kilowatt hour, which is less than standard estimates of the opportunity cost of involuntary rationing, and therefore within reason. The curtailment premium averaged 9 percent of original full spot prices, but the standard deviation of full spot prices almost tripled.^{5.1}

Figure 5.5.1

Full Spot Prices

Base Case versus "With Curtailment Premium" Case

Price Range (¢/kwh)	Hours per Year in this Range	
	Base Case	Curtailment Prem. Case
0 to 6	8608*	8608*
6 to 11.5	162	77
11.5 to 20	0	42
20 to 40	0	18
40 to 72	0	25
Mean price for year	2.63 ¢/kwh	2.86 ¢/kwh
Std. Deviation of price	1.35 ¢/kwh	3.86 ¢/kwh

* = See Figure 5.5.1 for detailed price duration curve in this range

Results

The procedures used to calculate prices under non spot rates A through E, and to model customer behavior for the simulated utility, are identical to those described earlier for the base case.

The most important results of these calculations are shown in Figure 5.5.2. (See Figures 5.3.3 and 5.3.4 for the complete base case results.) Results for the composite shutdown customer are not shown, since they were almost unchanged from the base case. The standby generator customer is shown for the first time. It would never turn on the generator except under full spot pricing, since prices under the

Figure 5.5.2
Results for "Spot Prices with Curtailment Premium" Cases
(Thousands of dollars per year; base case results in parentheses)

<u>Rate</u>	<u>Simple Storage Customer</u>		
	<u>Total Social Cost</u>	<u>Relative Social Value</u>	<u>Ex Ante Expected Subsidy</u>
A (Annual flat)	1257 (1155)	0 (0)	0 (0)
B (Annual t-o-u)	1256 (1152)	2 (3)	20 (16)
C (Monthly flat)	1257 (1155)	0 (0)	1 (0)
D (Monthly t-o-u)	1255 (1152)	2 (3)	23 (16)
E (Monthly 24)	1212 (1121)	45 (35)	1 (2)
F (Full spot)	1204 (1114)	53 (41)	0 (0)

<u>Rate</u>	<u>Discrete Rescheduling Customer</u>		
	<u>Total Social Cost</u>	<u>Relative Social Value</u>	<u>Ex Ante Expected Subsidy</u>
A (Annual flat)	480 (444)	- - -	61 (59)
B (Annual t-o-u)	481 (440)	-1 (4)	17 (18)
C (Monthly flat)	480 (444)	0 (0)	61 (59)
D (Monthly t-o-u)	481 (440)	-1 (4)	18 (18)
E (Monthly 24)	461 (435)	20 (9)	1 (1)
F (Full spot)	454 (431)	26 (13)	0 (0)

Composite Shutdown Customer

Not shown - results almost identical to base case. (See Figure 5.3.5 for base case.)

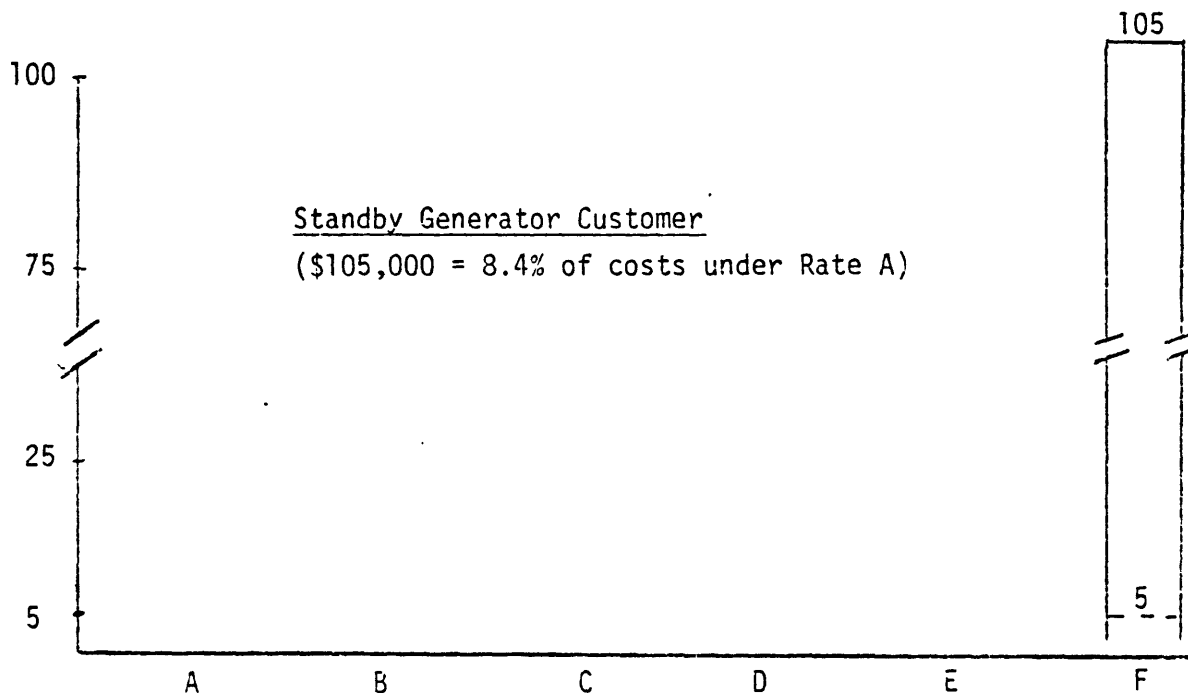
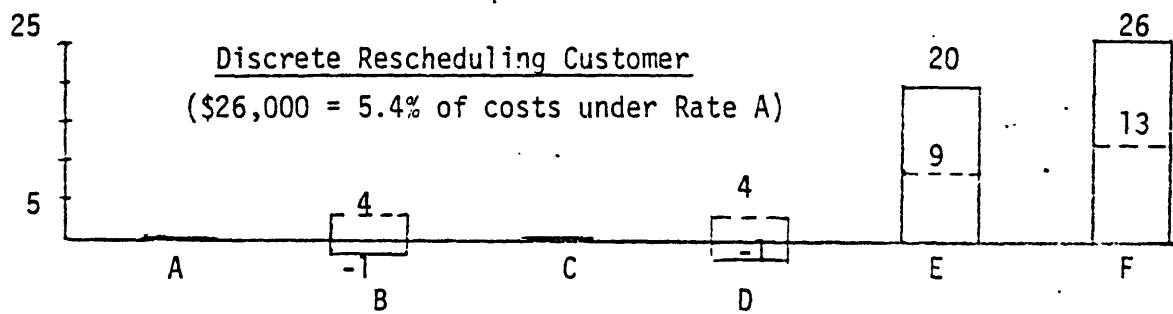
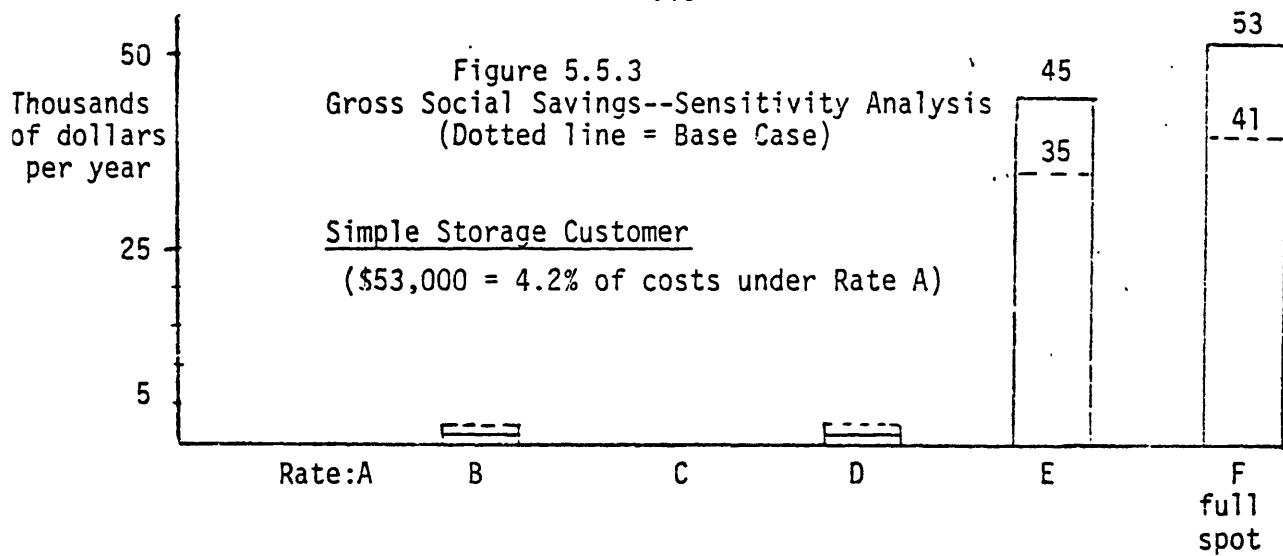
<u>Rate</u>	<u>Standby Generator Customer</u>		
	<u>Total Social Cost</u>	<u>Relative Social Value</u>	<u>Ex Ante Expected Subsidy</u>
A through E	1257 (1155)	0 (0)	Approx. 0(0)
F (Full Spot)	1152 (1154)	105 (1)	0 (0)

other rates never reach its marginal generating cost. Hence the other rates have uniformly high social and private costs for this customer.^{5.2}

The existence of the curtailment premium leaves unchanged or increases social costs (full spot prices) at all hours. Therefore social total annual costs are at least as high, for every customer and every rate. However, proper demand altering behavior is socially more valuable, and the losses caused by the incorrect incentives of Rate A are more costly. Therefore the social value of full spot prices increases. This is graphically portrayed in Figure 5.5.3. Rate E is close enough to full spot prices that its social value also increases, except for the standby generator customer.^{5.3}

The other rates do not do as well. The curtailment premium, which is only significant for a few hours a year, is badly approximated by the two-level time-of-use rates B and D. Hence customer behavior under these rates is the same in the curtailment premium case as it was in the base case. Because we are valuing that behavior at new prices, it happens that the social value of time-of-use rates falls. In fact for the discrete rescheduling customer, society is now worse off if the customer is on time-of-use rates than if it is on flat prices. Thus it is not always true that rates which are closer to full spot prices are better. Fortunately, the social loss appears to be quite small.^{5.4}

Because of the large absolute dollar increase in the social values of full spot prices between the two cases, full spot prices are more likely to be the socially preferred rate for customers in the "with premium" utility's territory than in the base case utility's territory. Because



subsidy levels are approximately unchanged, these additional benefits are passed on to customers, and customers would be more likely to self select full spot prices.

Value of Better Price Forecasts

Any rate which calculates prices in advance has some forecast error. These forecast errors determine the optimal interval between price recalculations; see Section 2.5. The better prices can be forecast, the less serious is a given interval between price recalculations. To study this, I simulated the effects of better one-month ahead forecasts for the base case utility (no curtailment premium).

Recall from Section 5.1 that spot price forecasts under rates C, D, and E were modeled by using a weighted average of the true spot price, and the spot price the previous month. All results presented so far put a weight of .4 on the true value, and .6 on the previous month. Here I show what happens with an extremely good forecast: a weight of .8 on the true value, and only .2 on the past month. Rates A, B, and F were left unchanged for this analysis.

For the storage customers, this has almost no effect on behavior or social value. Their behavior is determined by the relative ordering of prices each day, not their absolute levels. Thus flat and two level time of use (rates C and D) were unaffected. For rate E the forecast orderings were more nearly correct, but the effect was only to switch demands between hours of approximately the same full spot price. Thus for a real utility setting prices one month in advance for storage

customers, it is only important to predict the precise hour of system peak price if that hour's price is very different than neighboring hours' prices. If the utility has a wide flat peak, any hour is almost as good as any other.

For the composite shutdown customer, however, the absolute level of prices is critical, and better forecasts had more value. The social and private costs of this customer's behavior under rate A are \$878,000 per year. Figure 5.5.4 shows the social value of each rate (reduction in costs below \$878,000) as a function of the forecast quality.

Figure 5.5.4

Social Value of Better Spot Price Forecasts
(Percent of Costs Under Rate A)

	<u>Base Case</u>	<u>Superior Forecasts</u>
Rate C (Monthly flat)	0.1%	0.7%
Rate D (Monthly t-o-u)	2.0	3.5
Rate E (Monthly 24)	5.6	6.7
Rate F (Full Spot)	9.9	9.9

Although better forecasts did improve the value of these three rates for this customer, even 100 percent accurate forecasts could not do as well as full spot pricing. As shown in Figure 5.1.1, successive days and weeks can have very different full spot prices. But rates C, D, and E aggregate an entire month into two day types: weekdays and weekends. So even with 100 percent accurate forecasts, time aggregation error would cause losses. A rate which recalculated prices each week would do better.

Generalizations

Full spot prices show very different patterns for different U.S. utilities. We are now in a position to generalize about the relative merits of different rates, for the same customer type situated in different utilities. The key utility characteristics are the range, pattern, and predictability of spot price changes over the course of a year. These can be associated with the demand pattern and generating capital stock of the utility. For example:

- o Utilities with a lot of hydro power with storage will use it to level system lambda within a week. Therefore rates which are recalculated once a week or even once a month will capture most of the benefits of full spot pricing. The prices can be flat over the interval between recalculations. Once a month price recalculations (rate C) will be considerably better than once a year recalculations (rate A) because variable rainfall can cause considerable month to month change in system lambda.

- o If the utility is sometimes capacity limited, so that the curtailment premium is important, this is less true. Spot pricing will always have value for such utilities. This caveat also applies to all of the following. Also, marginal line losses will vary over a day, leading to some daily variation in full spot prices even for a heavily hydro utility.

- o Before 1973, the (short run) marginal costs of coal and oil generated electricity were approximately the same in much of the U.S., again implying relatively constant system lambda and little value of full spot pricing. (However some utilities were having difficulty meeting peak demands; for them spot pricing might still have been worthwhile for large customers.)

- o Utilities whose full spot prices are heavily determined by weather's effect on demand will benefit from prices recalculated no farther ahead than the weather can be accurately forecast. Full spot pricing will be more valuable in Florida (unpredictable weather) than in Arizona.

- o As the costs of real time communications and automated response fall, the breakeven customer size for full spot prices will also fall, on a utility with constant characteristics.

- o On most utilities there is significant within-day change in full spot prices. (Exceptions are the cases discussed above.) For these utilities two or three level time-of-use rates cause a significant time aggregation error, and the utilities should consider rates which change every hour during the morning load pickup period and other times when loads usually change rapidly. Even if prices are only recalculated every few months, such rates may justify the cost of more complex meters on large customers.

- o Utilities with less reliable supplies (heavy use of tie lines or a few large units) should use rates closer to full spot pricing.

- o Utilities with service territories which are unusually subject to losses during rolling blackouts should use rates closer to full spot pricing.

5.6 Nationwide Applicability of Spot Pricing

How much of U.S. electricity is consumed by customers which should be on for full spot prices or other hourly time varying prices? An accurate answer would require a utility-by-utility study. Here I present some relevant data which is available at an aggregate level. Depending on the screening method used and how tightly it is applied, from a few percent to one quarter of U.S. electrical use could profitably be on hourly prices.

We know from the case studies of this chapter that the optimal rate for a particular customer depends on interplay of three factors:

- o Customer type (discussed in Section 5.3)
- o Customer size (Section 5.4)
- o Characteristics of the local utility's full spot prices (Section 5.5)

A complete analysis would require a cross-tabulation of U.S. electricity use along these three dimensions.

The only data available shows electricity use distributed along one of these dimensions at a time. Specifically, I will discuss:

- o The detailed size distribution of large customers, for two utilities in the U.S.
- o The amount of electricity consumed in ten end uses which seem qualitatively suitable to spot pricing.

Size Distribution of Large Customers

For customers in the subject utility's territory, case studies suggest that hourly time varying rates (either rate E or rate F) are desirable for customers with suitable processes which have peak demands greater than approximately one megawatt. The subject utility currently has about 500 "general primary" customers. These are its largest customers; membership in this class is mandatory for customers with peak demands over 300 kilowatts. Customers in this rate class use about 5 billion kilowatt hours a year of energy, which is 31 percent of the utility's total sales. From data collected for another project, I estimated the distribution of customer size which is shown in the first two columns of Figure 5.6.1.

The best available customer size index data was peak kilowatt demand during 1980. The figure sorts customers by peak demand, and shows in the first column how many customers are in each size class. The second column shows what fraction of total energy use for the class was used by customers in that size class. For example, customers over 5 MW used 43 percent of all general primary electric energy, which is 13 percent of the utility's total sales. Thus a very small percentage of the customers consume a significant fraction of the utility's total sales.

Figure 5.6.1

Size Distribution of Customers

Customer Size	Subject Utility		Alternate Utility		
	Number of Customers	kwh used by customers this size	Number of Customers	kwh used by customers this size	
0.3 to 2 MW	65%	28%	61%	25%	
2 to 5 MW	22%	29%	28%	32%	
5 to 10 MW	8%	22%	7%	19%	
over 10 MW	4%	21%	4%	23%	
<u>Sample Group Only--</u>	Total (percent)	100%	100%	100%	
	Total (absolute)	538	5.0 x 10 ⁹ kwh	123	1.4 x 10 ⁹ kwh
<u>Entire Rate Class</u>	Total (absolute)	538	5.0 x 10 ⁹ kwh	450	3.4 x 10 ⁹ kwh
<u>Entire Utility</u>	0 to ∞ MW (Rate Class/Utility)	800,000	16.2 x 10 ⁹ kwh	770,000	12 x 10 ⁹ kwh
		.07%	31%	.06%	28%

Notes:

1. Data from the two utilities was derived from different sources, is for different years, and has different qualifications. The similarity between the two data sets is quite surprising.
2. The data on the subject utility is from hour-by-hour records of 1980 demand, by all of its customers with peak demands of 300 kw or greater. As shown, these customers used about 31 percent of the utility's total sales for 1980.^{6.0}
3. The "alternate utility" data is from a survey discussed by Pickel [1982]. It applies to a single year, approximately 1975. The survey was sent to customers with peak demands over 500 kw, rather than the 300 kw cutoff in the data for the first utility. The survey did not cover all customers, as shown in the third and fourth rows from the bottom. Therefore the size distribution of customers is less accurate for this utility.

Most customers over 5 MW in peak demand have an average demand of 2 or more MW, and have several large electricity using processes. This combined with the results of Section 5.4 suggest that most of this utility's customers over 5 MW peak belong on hourly time varying prices. A substantial fraction of the customers between 2 and 5 MW may also be suitable. Thus based on size alone, up to 20 percent of this utility's electricity use should go on full spot prices, corresponding to 190 individual customers.^{6.1} Ten percent is a conservative estimate.

The second half of Figure 5.6.1 shows comparable size data for another U.S. utility in about 1975. This data is based on a partial survey of customers over 500 kw peak demands [Pickel, 1982]. Also this data was for an earlier year, and used a higher size cutoff, making direct comparison with the subject utility impossible. Nonetheless it again suggests that over ten percent of this utility's total sales went to customers over 5 MW.^{6.2}

Electricity Use by Suitable Processes

The other way to estimate how much electrical load is a good candidate for spot pricing is to look at the nature of the processes which use electricity. Chapter 4 discussed some manufacturing and nonmanufacturing uses of electricity which appear to be good candidates for spot pricing because they have low shutdown thresholds, good storage/rescheduling possibilities, or both.^{6.3}

The indicated SIC's are shown in Figure 5.6.3, along with rough estimates of their electricity use and other relevant information. Total U.S. electricity use and U.S. manufacturing electricity are shown in the first two rows, for comparison. All data labeled "ASM" is for 1979. Other data is for other years, and is based on indirect calculations by various authors, hence is less reliable. For example, the commercial air conditioning data is for 1975. In addition, only chilled water air conditioning is well-suited to storage, and I have no data on how much commercial air conditioning uses chilled water.^{6.4} For this preliminary survey I have used a figure of 10%; therefore the column labeled "subuse" shows only 16 billion kwh. Similarly, for hospitals the relevant number is the amount of standby generation available; I used a conservative figure of 30% of the hospitals' average load.^{6.5, 6.6, 6.7}

Columns 6 through 10 of Figure 5.6.3 give information about how customers of each type would respond to spot prices.

The sixth column shows the nationwide average ratio of electricity use to value added for the SIC. A high ratio implies an electricity intensive process, and implies that if the spot price were to rise significantly, firms in this SIC would shut down by curtailing their total output, thereby increasing their profits. (See Section 4.3.) The reciprocal of this ratio, shown in the next column, shows how much spot prices would have to rise above their

Figure 5.6.3

Electricity Use in Suitable Processes

SIC	De- scrip- tion	Source	Elec- trical Purchase 10 ⁹ kwh	Appro- priate use	1979	Impli- cit shut- down premium ¢/kwh	Sui- ta- bility to shut- down entire plant	Sui- tabil- ity to stor- age
					Total kwh per \$ of value added			
Total U.S.		EEI	2,079	--	X	X	very low	--
20-39	All Mfg	ASM	682.4	--	1.0	100	very low	--
2621	Paper- mills	ASM	21.8	20	4.9	20	medium	medium
2631	Paper- board	ASM	11.6	10	6.2	16	medium	medium
2812	Alka- lies & Chlorine	ASM	10.8	10	22	4.5	high	low
2813	Indus. Gases	ASM	13.7	13	16.5	6.1	high	low
3241	Cement	ASM	10.3	7	4.7	21	low	high
3313	Electro- furnaces	ASM	8.7	8	20.3	5	high	medium
--	Agric. irrigtn.	IEUDB	19.2	19	X	X	low	high
461	Oil pipe- lines	Hooker	~10	10	X	X	low	medium
4941	Municpl. water	Text	~10	10	X	X	low	high

FIGURE 5.6.3 (continued)

<u>SIC</u>	<u>De- scrip- tion</u>	<u>Source</u>	<u>Elec trical Purchase 10⁹kwh</u>	<u>Appro- priate use</u>	<u>1979 Total kwh per \$ of value added</u>	<u>Impli- cit shut- down premium ¢/kwh</u>	<u>Suita- bility to shut- down entire plant</u>	<u>Suit- abil- ity to stor- age</u>
	Commercial air condition- ing--water	J&J	~160	16	X		low	medium
<u>806</u>	<u>Hospital standby generator</u>	<u>J&J</u>	<u>52</u>	<u>15</u>	<u>X</u>	<u>5</u>	<u>high</u>	<u>low</u>
<u>TOTAL</u>				<u>138</u>				

X = No data

Sources: EEI = Edison Electric Institute. Data is for 1979.
 ASM = Annual Survey of Manufactures for 1979.
 IEUDB = Industrial Energy Use Data Book [Oak Ridge Affiliated
 Universities, 1980]. Data is for 1974.
 Hooker = Calculations based on Hooker [1981].
 J&J = Jackson & Johnson. Data is for 1975, not 1979.

1979 average level to induce such shutdown. For example, for an "average" industrial gases plant (SIC 2813), spot prices 6 cents per kwh above the 1979 average price would cause the plant to shut down, even though this means reducing its production. (The 1979 average price for this SIC was 2.5 cents per kwh.)^{6.8}

The next to last column of the table uses this ratio to classify the SIC's suitability to this drastic form of shutting down. Shutdown costs below 8 cents per kwh are classified as "high suitability"; 8 to 20 cents as "medium". Of course a more likely and lower cost approach to shutting down is to shut down only the production of the most electricity intensive products within a plant or SIC. The aggregated data used here does not permit estimating the price thresholds for such response.

Finally, the last column indicates the likelihood of pure storage behavior by firms within the SIC. Pure storage behavior means producing the same amount of final product, but shifting the time of use of electricity by hours or days in response to spot prices. This is possible for plants with electricity intensive stages which produce storable intermediate products, as discussed in Chapter 4. For example, cement plants use roughly two-thirds of their electricity for grinding raw materials and semifinished product. These can be stored for hours or days until needed, without interrupting final production. [Gordian Associates, 1980, p. IV-9].

As modeled in Chapter 4 and this Chapter's case studies, the amount of such storage behavior depends on the plant's capital stock. Some cement plants might have very "tight" designs, with peak

grinding capacity little higher than the average requirement. Such plants would require new investments in order to take advantage of spot prices. But in many other industries, loads normally operate well below 100 percent of maximum capacity. Agricultural irrigation systems, municipal water systems, and air conditioners are examples. Oil pipelines also operate below full rated capacity for most of their economic lives. [Hooker, 1981; Gordian Associates, 1980]

In order to receive a "high" rating in the final column of Figure 5.6.3, a process must satisfy at least two of the following three criteria:

- o Electricity intensive intermediate product which can be stored easily and with low losses.
- o No capital investment needed for storage; the product is normally stored anyway.
- o The process is normally operated at less than 100 percent of rated output, so that storage can be charged up. Marginal operating costs are a nonincreasing function of the operating rate, so there is no inherent technical reason to smooth electricity use.

Satisfying only one criterion completely and one moderately resulted in a "medium" rating.

Conclusion

Total suitable electricity use in these eleven categories was approximately 140 billion kwh, or 6.7 percent of total 1979 electrical use.^{6.9} My rough categorization indicates that it is divided about equally between shutdown response and storage response. Of course if full spot pricing with a curtailment premium were ever implemented,

and spot prices similar to those in the sensitivity analysis case of Section 5.5 resulted, much more shutdown behavior would take place.

The customer size oriented approach discussed in the first part of the section gives very different results than the SIC oriented approach of the second part. Using a size cutoff of 5 MW suggested that over ten percent of U.S. electricity use is a suitable candidate for hourly time varying prices. Many, perhaps most, large customers are not in any of the SIC's evaluated here. However, individual electricity intensive processes within many other large customers may be good candidates. Analysis of processes within plants would be needed to measure these.

A triple screening by customer size, customer type, and utility characteristics would clearly lead to still lower estimates of how much demand is suitable for spot pricing. To be accurate such an analysis would have to use the customer lists of individual utilities. Such analysis might turn up other suitable customers, such as non-hospitals with standby generators. Small power producers, many of which are good candidates for full spot pricing, should also be identified.

Even if it is not optimal to put a customer on full spot pricing, some rates other than flat prices may be appropriate. In the case studies, customers over several hundred kilowatts belonged on two level time-of-use rates or something more sophisticated. Customers of this size apparently use at least 30 percent of electricity sold in the U.S.

5.7 Conclusions

The numerical results of the case studies in this chapter are suggestive but not close to definitive. The methods used to generate them are limited by many simplifying assumptions. Furthermore results will change from utility to utility, as shown in Section 5.5.

The major simplifying assumptions of this chapter include:

- o The prices generated under the different rates are only approximate, as discussed in Section 5.1. System lambda was used as a proxy for full spot prices, and will underestimate the variability of actual full spot prices because it ignores spatial and demand curtailment components. I also use approximate "forecasting" methods to simulate forecast errors for the non-spot rates.
- o All "full" spot prices were hourly averages. If this utility has significant variation in system lambda within an hour, prices which change more often would give higher gross benefits.
- o Other rates could have been evaluated, such as daily or weekly instead of monthly price recalculations.
- o The capital stock of each customer is treated as exogenous. But in fact under spot pricing customers might install more capital to take advantage of price changes, increasing the long run value of spot pricing.

- o The underlying schedule of each plant is treated as fixed rather than exogenous.^{7.1} Under some rates it may be profitable to shift some operations to the night. Ignoring this understates the benefits of time varying prices but not the incremental benefits of full spot prices.
 - o The cases ignore response modes which require more than one day's warning. For example, some of the sites visited could reschedule production within a week as well as within a day. This would be more profitable under full spot pricing than under other rates, but the difference depends on how accurately prices could be forecast several days ahead.
 - o Only selected and stylized customer processes were modeled. Other processes would show different responses. For example assembly line operations will not be very responsive to electricity prices, since their high labor use implies high value added per kilowatt hour. Only in the rare case that workers could be used for other productive tasks while the line was shut down would a customer consider temporary shutdowns in midday.
- Subject to these limitations, the results of the case studies still suggest several points. Most of the qualitative points which follow can be explained on theoretical grounds, and therefore should generalize to other customers on similar utilities.

- o For storage/rescheduling type customers, rate E (monthly recalculation of hourly prices) gives more than two thirds of the gross benefit of full spot prices, with considerably lower transactions costs.
- o Time aggregation error appears to be a major source of social loss. The two level time-of-use rates give less than half the gross social benefits of rate E with its hourly price changes. Yet the mandatory incremental transactions costs of rate E are low for customers who already have the required hourly metering equipment. Hence rate E should automatically be considered for any customer with hourly metering, which in many utility jurisdictions includes all the largest customers.
- o Full spot prices can in extreme cases give up to ten times the gross social benefits of conventional two-level time-of-use rates.
- o Of the customers studied, shutdown customers with the proper shutdown level seem to benefit the most from time varying prices, especially full spot prices. The gross value of full spot pricing for the shutdown customer modeled was roughly ten percent of its total energy costs, in the base case utility region.^{7.2} Some pure storage customers with several hours of storage would also have benefits this large.
- o The difference between the value of spot prices and the value of predetermined prices which have the same amount of time aggregation depends on how well prices can be forecast, and on

the customer. Shutdown customers show more improvement from spot prices (rate F compared with rate E) than do storage customers.

- o Cross subsidies under various rates will indeed lead to socially suboptimal self assignment, as was discussed in Section 3.5. However for the customers modeled here, only flat prices led to really large cross subsidies. Therefore there would be little mis-assignment by allowing customers to select among different time varying rates.
- o In order to assign different customers to different rates efficiently, it is necessary to know more about them than their size. The price at which they will shut down, the number of hours of storage capacity, and their normal operating schedule (weekdays or all week) are useful statistics for this purpose. It also makes a difference whether the customer has automated control of its electrical loads, since this will reduce its transactions costs.
- o Investment incentives are strongly influenced by the rate the customer is on. In all practical cases considered, full spot prices increase the social value of a given investment, and also increase the profits which would be earned by its owner. Again this is particularly true for "shutdown" investments, such as self generation equipment.

FOOTNOTES TO CHAPTER 5

1.1 Approximately 15 percent of the days over this period were missing. The missing days were "selected" randomly or close to randomly by the utility. I filled in the missing days by averaging the two adjacent days to approximate the missing day. For weekdays, only other non-holiday weekdays were used. For Saturdays, the closest Saturdays were used. For Sundays, the closest Sundays were used. One legal holiday was missing and was corrected assuming it was like a Sunday. The use of this averaging procedure probably biased the data slightly toward less variability in system lambda, and therefore toward less benefit from full spot prices.

1.2 An alternative to using historical system lambda would be to synthesize values by simulating the utility's own short run supply curve, and using historical demand. This is a fairly standard technique for electric utility modeling. But the subject utility purchased a lot of energy from neighbors during 1980, and this procedure would have given rather inaccurate results. Therefore I chose to use the more realistic actual historical data on system lambda.

1.3 This is a conservative assumption, as it will understate the true variability and benefits of full spot prices.

1.4 Actual full spot prices probably change more than once an hour, but only hourly system lambda was available. Again this assumption will understate the benefits of true full spot prices.

1.5 Except that rates B and C cannot be ordered, since rate C has a shorter interval between price recalculations but a longer interval between price changes.

1.6 The total cost of electricity includes a time-of-use energy charge, a time-of-use unratcheted demand charge, and a negligible fuel adjustment charge. The demand charge changes seasonally but not with time of day. The energy charges do not change seasonally.

1.7 1980 was a leap year, 8784 hours long.

1.8 Using net revenue under full spot prices as the "base point" for comparison is for convenience and has little effect on the results. Section 3.5 showed that the subsidy formula requires measuring behavior under the non-spot rate only. Cross subsidies for case study customers will be calculated below.

1.9A lower bound on the spot price forecast variance is the variance of fuel price forecasts. The nominal dollar price of residual fuel oil increased by 32 percent from December 1979 to December 1980. The corresponding national average increases for natural gas and coal were 23 percent and 7 percent respectively. [Monthly Energy Review, 1981] Some of this increase might have been anticipated, of course, and these year to year changes do not directly imply any particular variance. Nonetheless this is further evidence that the 5 percent error is conservative.

1.10 If prices under the other rate are $p'(t)$, the last column shows the standard deviation over the year of $p'(t) - p^*(t)$. The second column shows the standard deviation of $p'(t)$.

2.1 For simplicity I assume that all four of these customers have deterministic schedules which are also constant from one day to the next. All except the discrete rescheduling customer are assumed to operate around the clock; sensitivity analysis on this will be mentioned later. I did not model any stochastic underlying demands, such as weather sensitive demands. Presumably they would not show any fundamentally different results, except that only rates with very frequent price recalculations could eliminate the subsidies to such customers. That is, the zero subsidy property which we will see later for rate E would not hold for customers with stochastic demands which are positively correlated with spot prices. See Section 3.5.

2.2 At the level of individual machines, capital stock is often not "fully utilized," i.e. not operating every minute of a shift. A plant may have only one or two of a particular piece of equipment, and they will be sized to meet maximum rather than average loads. The machines may also have different production rates for different product variants, and be sized for the worst case. In a world with uncertainty, demand variation, learning by doing, and scale economies, it is not necessarily "inefficient" to have equipment which sits idle part of most labor shifts.

Also, except in a few industries, it is rare to operate a plant around the clock. Thus some rescheduling is generally possible if the electricity savings outweigh the additional labor costs. I do not model that here, however.

2.3 Under full spot pricing, prices will not be known with certainty 24 hours in advance. I assume perfect one-day ahead forecasting as an approximation. On most systems, weather and demand can be forecast accurately this far ahead and price forecasts will therefore be correct except on days with unplanned unit or transmission line outages.

2.4 Many such thermal processes today will not have a full hour of storage capacity. In such cases the benefits are reduced in exact proportion to the actual number of hours of storage. Also the assumption that recharge is possible within one hour implies a maximum heating or cooling rate which is twice the average demand. This is reasonable for some processes but not for commercial building space conditioning on days with particularly extreme weather.

2.5 For thermal processes, losses will encourage waiting as long as possible before filling storage, and discharging it as soon as possible. I modeled this by assuming that if prices are at a constant low level for several hours, storage is filled in the last of those hours, and conversely it is discharged at the beginning of a period of constant high prices. With two-level time-of-use prices (rates B and D) this makes a big difference. Customers fill storage from 7 AM to 8 AM, and empty it from 8 AM to 9 AM after the price jump. This behavior gives relatively little social benefit, since the full spot price normally does not climb much over that interval. But it is indeed profit maximizing.

2.6 Most so-called "run-of-river" hydro systems actually can control how much water flows through the turbine, and therefore can store water. In New England several hours to one day of storage capacity is usual. Personal communication, F. Pickel, May 20, 1982.

2.7 This can be expressed in the notation of Chapter 4 as :

$$\sum_t X(t) = \text{constant}$$

$$X(t) \leq X_{\max}$$

2.8 This site visit was conducted by R. Tabors and M. Caramanis, not by the author.

2.9 The firm currently finds it economical to operate ten hours per day instead of eight because running its melter at less than maximum significantly reduces the firm's demand charge. Many systems with storage are currently operated to reduce demand charges rather than energy charges. None of the rates considered here have a demand charge.

Under a new rate it might be optimal for this customer to be open either more or less than 10 hours per day. To assess this would require knowing its labor cost structure and its opportunities to put furnace workers on other tasks during a mid-day shutdown.

2.10 The relative efficiency of electricity and other fuels is process specific. Because electricity can be applied more precisely and has no thermal loss via flue gases, and because natural gas combustion efficiency alone is below .9, electricity is almost always more efficient. If product cleanliness is a major constraint the difference can be quite large since indirect application, perhaps via steam, may be necessary. To cite one example where product cleanliness is not a constraint, Williams [1981] gives an efficiency of 28 percent for a conventional glass pellet melter and 61 percent for a melter with a preheating furnace. Electric heating would not be 100 percent efficient, but it would be higher than 61 percent. My thanks to G. Russo for help on this subject.

2.11 The heat rate and maintenance costs are from Pickel [1982]. The fuel cost is the average 1980 delivered price of distillate fuel, from Monthly Energy Review [1981].

3.1 For the storage customer, the only energy costs are for electricity. For the shutdown customer, this column measures electricity plus fossil fuel consumption.

3.2 These results were verified by side analysis and sensitivity analysis on similar customers, not shown. Approximately 20 variants of these customers were modeled.

3.3 Actual benefits of this rates may be larger than shown here, because I do not model permanent shifts in daily or weekly demand patterns due to predetermined time-of-use rates. For example some firms may change workers to a night shift, which would increase the social benefits of all time varying rates.

3.4 In fact not all storage customers will follow this pattern exactly. For example the discrete rescheduling customers might schedule the "break" in the middle of the high price period, for labor reasons.

3.5 Rate E would not remove the time aggregation error subsidy for a customer which operated mainly on certain days of the week. For example firms which operate on Saturday but not on Sunday would still have a slight positive subsidy, as full spot prices tend to be higher on Saturday.

3.6 Of course extreme customers could be constructed which would have large subsidies even under two-level time-of-use rates.

3.7 This relationship holds for "normal" investment, but one can construct pathological counter-examples, in which an investment has greater social value under non-spot prices. An intuitive discussion goes as follows. Any new investment increases the customer's behavioral options. If under full spot pricing, it will select new behavior which has the largest possible social benefit, whereas if it is on another rate it might exploit the investment in a socially less productive way. Thus the investment has a higher social value under full spot prices.

Nonetheless under non spot prices the new investment might change the customer's entire behavior (not just the behavior associated with the new equipment), and do so in a way which increases social value. For example consider the simple storage customer under two level time-of-use prices. Suppose that with only a small amount of storage, it discharges the storage from 8 AM to 9 AM, which reduces the social value of the storage. Suppose it then invests in much more storage capacity. This might now make it worthwhile for the customer to reprogram the way all storage is used, and do no discharging between 8 and 9 AM. (For example if less labor is used while storage is discharging, the labor union might require that discharging be done in the afternoon, as a condition of not fighting the investment. Under two-level time-of-use rates, the customer is indifferent to when discharging takes place, so would agree to the union request.) This would increase the social value of the investment under this rate, perhaps enough to make it higher than the social value under full spot prices.

This is obviously a stilted counter example. In all of the "normal" investments I have considered, social savings under full spot prices are at least as high as under any other rate.

3.8 If the customer's underlying production technology or cost function is nonlinear, then the evaluation of a plant expansion must be modified somewhat. The central point which follows will still hold.

3.9 For this special case of an expansion of an entire plant which scales up electricity use at all hours by the same amount, these results always hold.

3.10 Figure 5.3.8 is based on sensitivity analysis of the discrete rescheduling customer.

3.11 Figures 5.3.6 and 5.3.7 are based on the estimates of transactions costs in Section 5.2. Since those estimates were approximate, the crossover points and social loss estimates are also.

3.12 Except for customers with weather correlated demands.

4.1 For example most of the customers on the San Diego "coincident demand" rate elect not to pay for and receive the real time "pseudo price" signal made available by the utility. [Gorzelnik 1979]

4.2 For example if the load is already controlled by a process control computer, the initial cost might be reprogramming the computer; the monthly cost would be feeding in the vector of 48 prices for the month. Alternately it might be cheaper to simply decide what to do each month by hand, then directly instruct the computer to turn the load on and off at specific times each day.

4.3 A lower cost alternative is to get a vector of 24 hourly prices, recalculated once a day. For such a rate communications costs would be much lower than under full spot pricing. Newspapers or a daily telephone call would suffice.

4.4 Price forecasts can be calculated by the customer, but more likely is that the utility will transmit them at the same time as the current price.

4.5 If next hour's price has an expected value equal to the current price, behavior should be based on the following hour's expected price.

4.6 Under rate E, this customer would have to take action an average of about five times a day, or one fifth as often as under full spot pricing.

4.7 A real time information processing system would signal the plant operator when he had to take an action in response to a price change, and tell him what action to take.

4.8 Of course if the customer is on full spot pricing, it gets no cross-subsidy, even if it chooses not to respond to the current price. The gross social benefits of full spot prices, however, are changed to those of a less sophisticated rate.

5.1 All of the changes were on weekdays; weekend and holiday full spot prices were the same as before. Most but not all of the changes were during peak hours (8 a.m. to 8 p.m.).

5.2 The value of the standby generator is \$105,000 per year if the customer is under full spot prices. By comparison Pickel [1982] estimates the capital cost of such a generator to be \$2.7 million. Thus spot prices are not high enough here to pay for a diesel generator, unless it is also needed for emergency use.

5.3 Since the composite shutdown customer is almost always shut down under rate A, E, or F during the hours which are affected by the curtailment premium, the change in electricity prices is irrelevant to it. This explains why it is unchanged from the base case.

5.4 This loss happens because of a somewhat complex interaction between behavior and full spot prices. I model this customer under flat prices as operating steadily at 80 percent of its maximum possible operating rate, from 6 a.m. to 4 p.m. Under rates B or D it instead shuts down from 8 a.m. to 10 a.m., and operates at 100 percent the other eight hours. It happens that the average 1980 full spot price from 8 a.m. to 10 a.m. is, in the "with curtailment premium case", quite a bit lower than the price from 10 a.m. to 4 p.m. Hence the customer under rate B or D is saving electricity of comparatively little social value, and using more of the socially expensive electricity.

6.0 The hour-by-hour records were somewhat incomplete for about half of the 538 customers. Gaps were filled by interpolation. A check against a control total for the class showed a 5 percent error in total kwh consumption. The source of the first two columns is a customer-by-customer summary of the 538 customers, made from the hour-by-hour records.

6.1 Recall from the discussion in Section 5.4 that if a customer is put on full spot prices (rate F) and given the option of whether or not to pay to receive real time prices, it will effectively assign itself to rate E by not receiving prices, if that is the socially optimal behavior. Thus the utility can initially assign customers to full spot pricing, and let them in effect choose between rates E and F.

6.2 The only national data on customer size is very aggregated. Edison Electric Institute publishes data on total sales and sales to "industrial customers". Unfortunately the definition of "industrial" is based on Standard Industrial Classification (SIC) codes for some utilities. The result is that some customers below 300 kw are included, but some large non-industrial customers are excluded. (For the subject utility's 538 general primary customers, 218 had non industrial SIC codes.) Thus this national data will underestimate the national concentration of electricity demand. Figure 5.6.2 shows the EEI data for the total U.S. and for the East North Central region, which is the most concentrated region in the EEI statistics.

FIGURE 5.6.2
Comparisons of Estimated Energy Use By Large Customers

	<u>"Large" Customers</u>		<u>Energy Use</u>		<u>Average load per customer</u>
	<u>Number</u>	<u>Pct. of Total</u>	<u>Billion kwh</u>	<u>Pct. of Total Use</u>	
U.S., 1980, "Industrial"	485,200	0.53%	791	38%	180 kw
East North Central Region, 1980	16,710	0.39%	156	42%	270 kw
Subject utility, 1980, "General Primary"	538	0.07%	5.0	31%	1,060 kw
Alternate utility, approx. 1975	450	0.06%	3.4	28%	862 kw
U.S. approx. 1978, " _ 500 kw peak"	200,000	0.2 %	?	39%	approx. 450 kw

NOTE: See text for cautions. The rows are based on different size cutoffs.

Sources: Rows 1 and 2 from Edison Electric Institute [1981]. Rows 3 and 4 from Figure 5.6.1. Row 5 from D. Berkowitz (Westinghouse) presentation, May 1980.

Clearly the EEI "Industrial" category includes some customers smaller than the large rate class of either utility. This is confirmed by the Westinghouse estimate, shown on the bottom row, that 39 percent of electrical energy sales are to customers over 500 kw.

6.3 The best available information on how different SIC's use electricity is contained in reports on the potential for time-of-use pricing for different SIC's. One caution is needed before applying these reports to spot pricing. Permanent rescheduling of some labor may be feasible under time-of-use pricing. Customers which responded to time-of-use prices by rescheduling large numbers of workers would show little incremental benefit from spot pricing, since in most situations day-by-day labor rescheduling is impractically expensive, unless it can be done selectively for part of the work force. I therefore used judgement to decide whether the response mode described was flexible enough to give a good response to full spot prices.

6.4 Other air conditioning can use the thermal mass of the building for energy storage, but chilled water air conditioners can use a water tank plus the circulating water for storage.

6.5 Emergency generators are required for hospital accreditation; thus most U.S. hospitals will have them. They must be sized to supply all operating rooms and patient care areas, which may in fact mean a size equal to their average load.
Source: Personal communication, Betsy Boehner, Director of Project Review, Massachusetts Central Health Planning Agency.

6.6 The estimate of electrical use for municipal water pumping is based on extrapolation from the subject utility, which supports an area of about 2 million people. Eight of its general primary customers are SIC 4941. They used 89,700,000 kwh in 1980. Scaling this by a factor of 110 (for the U.S. population of 220 million) gives 10 billion kwh. This is obviously very approximate.

6.7 For the manufacturing SIC's, other than cement, I did not attempt to estimate an "appropriate use" scaling factor. Although not all electrical use in these SIC's could be controlled, the Annual Survey's procedure of assigning plants to a single SIC probably means that nearby SIC codes also contain some electrical use of the indicated type. This is especially true for SIC 3313, electrometallurgical products; much electricity used elsewhere in SIC 331 is probably for electrical furnaces.

6.8 The calculation of the shutdown point for hospital standby generators (part of SIC 806) was done differently. From the calculations in Section 5.2, the threshold for a standby diesel generator was about 8 cents/kwh, or about 5¢/kwh above the average price in 1979.

6.9 These numbers are approximate for the following reasons:

- o Conservative but subjective estimates were used to go from electricity purchases to appropriate use.
- o Some data was for 1974 or 1975, not 1979.
- o Industries and processes included were selected by quick screening, rather than individual considerations of all possible candidates.
- o The electricity purchase figures are estimates. The Annual Survey of Manufactures puts each plant into a single SIC code, even if it engages in multiple activities. The nonmanufacturing numbers were estimated by a variety of indirect methods.

7.1 For the discrete rescheduling customer, narrow variation was allowed by shifting the two hour "window" back and forth. But the operating hours were assumed to always be from 6 AM to 4 PM.

7.2 The key to the benefits of spot pricing for shutdown customers appears to be that the shutdown point be in the midrange of full spot prices during the hours that the customer operates. For example a customer identical to the "composite shutdown" customer but only open from 8 AM to 4 PM on weekdays has essentially no benefit from full spot prices, since between those hours they are almost always above its shutdown point. Any marginal cost based rate would lead the customer to not use electricity anyway.

On the other hand a customer with a high shutdown price, for example the standby generation customer, gets most of its benefit from spot pricing during the weekdays.

BIBLIOGRAPHY

- Acton, Jan Paul and Bridger Mitchell [1980] "Evaluating Time-of-Day Electricity Rates for Residential Customers", in Mitchell and Kleindorfer, eds., pp 247-273.
- Acton, J. P., E. H. Gelbard, J. R. Hosek, D. J. McKay [1980] British Industrial Response to the Peak Load Pricing of Electricity, Rand Corporation Report R-2508-DOE/DWP.
- Agnew, Carson E. [1973] "The Dynamic Control of Congestion-Prone Systems Through Pricing", Ph.D. dissertation, Stanford University. report R-2508-DOE/DWP, Santa Monica, California, February, 1980.
- Agnew, Carson E. [1977] "The Theory of Congestion Tolls", Journal of Regional Science, Vol. 17, No. 3, pp 381-393.
- Barro, R. J. [1972] "A Theory of Monopolistic Price Adjustment", Review of Economic Studies, vol. 39, pp 17-26.
- Barro, R. J. [1977] "Long Term Contracting, Sticky Prices, and Monetary Policy", Journal of Monetary Economics, vol 3, pp 305-316.
- Berg, Sanford V. [1981] "Load Management: Rationing Versus Peak Load Pricing", The Energy Journal, Vol. 2, No. 1, pp 89-98.
- Black, Fischer [1976] "The Pricing of Commodity Contracts", Journal of Financial Economics, Vol. 3, pp 167-179.
- Bohn, Roger E. [1980] "Industrial Response to Spot Electricity Prices: Some Empirical Evidence", MIT Energy Laboratory Working Paper MIT-EL-80-016WP, February 1980.
- Bohn, Roger E. [1981] "A Theoretical Analysis of Customer Response to Rapidly Changing Electricity Prices", MIT Energy Laboratory Working Paper MIT-EL-081-001WP, August 1980, revised January 1981.
- Bohn, R. E., M. C. Caramanis, F. C. Schweppe [1981] "Optimal Spot Pricing of Electricity: Theory", MIT Energy Laboratory Working Paper MIT-EL-081-008WP.
- Bohn, R. E., R. D. Tabors, B. W. Golub, F. C. Schweppe [1982] "Deregulating the Electric Utility Industry", MIT Energy Laboratory Technical Report MIT-EL-82-003.
- Boiteux, M. [1949] "La tarification des demandes en point: application de la theorie de la vente au cout marginal", Revue Generale de l'electricite, Vol. 58, pp 321-340.

Brown, Jr., G. and M. B. Johnson [1969] "Public Utility Pricing and Output Under Risk", American Economic Review, Vol. 59, March 1969, pp 119-129.

Business Week [1981] "Labor Seeks Less", December 21, 1981, pp. 82-90.

Camm, Frank [1980] Industrial Use of Cogeneration Under Marginal Cost Electricity Pricing in Sweden, RAND report WD-827-ePRI.

Caramanis, M.C., R.E. Bohn, and F. C. Schweppe [1982] "Optimal Spot Pricing: Practise and Theory", paper 82 WM 115-4 forthcoming in IEEE Transactions on Power Apparatus and Systems.

Carlton, Dennis W. [1981] "The Disruptive Effect of Inflation on the Organization of Markets", Working Paper No. 7, Program in Law and Economics, University of Chicago, 1981.

Craven, John [1974] "Space-Time Pricing for Public Utilities", The Bell Journal of Economics and Management Science, Vol. 5, pp 187-195.

Crew, M. A. and P. R. Kleindorfer [1976] "Peak Load Pricing with a Diverse Technology", The Bell Journal of Economics, Spring 1976, pp 207-231.

Crew, M. A. and P. R. Kleindorfer [1979] Public-Utility Economics, St. Martins Press, NY.

Crew, M. A. and P. R. Kleindorfer [1980] "Public-Utility Regulation and Reliability with Applications to Electric Utilities", in M. A. Crew, 1980.

Crew, Michael A., editor [1980] Issues in Public-Utility Pricing and Regulation, Lexington Books, Lexington, Mass.

Dansby, Robert E. [1979] "Multi-Period Pricing with Stochastic Demand", Journal of Econometrics, Vol. 9, pp 223-237.

Dansby, Robert E. [1980] "Spatial Considerations in Public Utility Pricing", in Michael Crew, Issues in Public-Utility Pricing and Regulation, Lexington Books, Lexington, Mass.

Danzinger, Leif [1981] "Uncertain Inflation and the Optimal Policy of Price Adjustment", Working Paper 27-81, Foerder Institute for Economic Research, Tel Aviv University.

Eaton, D. C. and R. G. Lipsey [1978] "Freedom of Entry and the Existence of Pure Profits", The Economic Journal, Vol. 88, pp 455-469.

Ebasco Services Inc. [1977] Rate-making: Topic 5 and Illustrative Rates for Five Utilities, Electric Utility Rate Design Study, Electric Power Research Institute, June 1977.

Edison Electric Institute [1981] Statistical Year Book of the Electric Utility Industry/1980.

Elgerd, Oile I. [1971, 1982] Electric Energy Systems Theory: An Introduction, first edition, McGraw-Hill Book Company, New York, 1971. Second edition, 1982.

Elgerd, Oile I. [1977] Basic Electric Power Engineering, Addison-Wesley Publishing Company, Reading Massachusetts.

El-Hawary and G. S. Christensen [1979] Optimal Economic Operation of Electric Power Systems, Academic Press, New York.

Ellis, Randall P. [1981] "Electric Utility Pricing and Investment Decisions Under Uncertainty", PhD dissertation, Massachusetts Institute of Technology.

Fischer, Stanley [1977] "Long Term Contracting, Sticky Prices, and Monetary Policy: A Comment", Journal of Monetary Economics, vol. 3, pp 317-323.

Gellerson, Mark W. and Shawna P. Grosskopf [1980] "Public Utility Pricing, Investment, and Reliability Under Uncertainty: a Review", Public Finance Quarterly, Vol. 8, pp 477-492.

Gordian Associates Inc. [1980] "Industrial Response to Time-of-Day Pricing: A Technical and Economic Assessment of Specific Load Management Strategies", Electric Power Research Institute Final Report EPRI EA-1573.

Gordon, Richard L. [1975] U.S. Coal and the Electric Power Industry, Johns Hopkins University Press.

Gorzelnik, Eugene F. [1979] "T-O-U rates cut peak 3.5%, kWh by 1.3%", Electrical World, September 15, 1979, pp 138-139.

Gorzelnik, Eugene F. [1980] "Regulators Look at Interruptible Rates", Electrical World, September 15, 1980, p 109.

Gorzelnik, Eugene F. [1982] "Special Report: Utilities Embrace Direct Load Control", Electrical World, February 1982, pp 69-84.

Gray, Jo Anne [1978], "On Indexation and Contract Length", Journal of Political Economy, vol. 86, pp 1-18.

Grossman, Sanford J. [1981] "An Introduction to the Theory of Rational Expectations Under Asymmetric Information", Review of Economic Studies, Vol. 46, pp 541-559.

Hall, Robert E. [1981] "The Inefficiency of Marginal-Cost Pricing and the Apparent Rigidity of Prices, xeroxed, Stanford University.

Hey, John D. [1979] Uncertainty in Microeconomics, New York University Press.

Holthausen, Duncan M. [1979] "Hedging and the Competitive Firm Under Price Uncertainty", American Economic Review, pp989-995.

Hooker, John N. [1981] "Oil Pipeline Energy Consumption and Efficiency", ORNL-5697, Oak Ridge National Laboratory.

Johnson, Lynwood A. and D. C. Montgomery [1974] Operations Research in Production Planning, Scheduling and Inventory Control, John Wiley and Sons, New York.

Joskow, Paul [1974] "Inflation and Environmental Concern: Structural Change in the Process of Public Utility Regulation", Journal of Law and Economics, vol 17, pp 291-328.

Joskow, Paul L. [1976] "Contributions to the Theory of Marginal Cost Pricing", The Bell Journal of Economics, Vol. 7, No. 1 (Spring 1976), pp 197-206.

Kepner, John and Mo Reinbergs [1980] "Pricing Policies for Reliability and Investment in Electricity Supply as an Alternative to Traditional Reserve Margins and Shortage Cost Estimations", unpublished manuscript.

Klein, B., R. Crawford, A. Alchian [1978] "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process", The Journal of Law and Economics, vol 21, pp 297-326.

Koenker, Roger [1979] "Optimal Peak Load Pricing with Time-Additive Consumer Preferences", Journal of Econometrics, Vol. 9, pp 175-192.

Leibenstein, Harvey [1980] Beyond Economic Man, Harvard University Press, Cambridge, Mass.

Littlechild, Stephen C. [1972] "A State Preference Approach to Public Utility Pricing and Output Under Risk", The Bell Journal of Economics and Management Science, Vol. 3, pp 340-345.

Lodish, Leonard M. [1980] "Applied Dynamic Pricing and Production Models...", Journal of Marketing Research, pp 203-224.

Luh, Peter B., Yu-Chi Ho and Ramal Muralidharan [1982] "Load Adaptive Pricing: An Emerging Tool for Electric Utilities", IEEE Transactions on Automatic Control, April 1982.

Manichaikul, Y. and F. C. Schweppe [1980] "Physical/Economic Analysis of Industrial Demand", IEEE Transactions on Power Apparatus and Systems, Vol. PAS-99, No. 2, March/April 1980, pp 582-588.

Marchand, M. G. [1974] "Pricing Power Supplied on an Interruptible Basis", European Economic Review, pp 263-274.

Marino, Anthony M. [1978] "Peak Load Pricing in a Neoclassical Model with Bounds on Variable Input Utilization", The Bell Journal, Vol 9, pp 249-259.

Mitchell, Bridger M. [1980] "Alternative Measured-Service Structures for Local Telephone Service", in Michael A. Crew [1980], pp 107-123.

Mitchell, Bridger M., P.R. Kleindorfer [1980] Regulated Industries and Public Enterprise, D.C. Heath and Company, Lexington, Mass.

Mitchell, Bridger M., W. Manning and J.P. Acton [1979] Peak Load Pricing: European Lessons for U. S. Energy Policy, Ballinger Publishing Co., Cambridge, Mass.

Monthly Energy Review, published by the U.S. Department of Energy, Energy Information Administration.

Morgan, M. Granger and Sarosh N. Talukdar [1979] "Electric Power Load Management: Some Technical, Economic, Regulatory and Social Issues", Proceedings of the IEEE, Vol. 67, No. 2, February, 1979, pp 241-313.

Mussa, Michael [1981] "Sticky Prices and Disequilibrium Adjustment in a Rational Model of the Inflationary Process", American Economic Review, Vol. 71, pp 1020-1027.

Nguyen, D. T. [1976] "The Problem of Peak Loads and Inventories", Bell Journal of Economics, Vol. 7, Spring 1976, pp 197-206.

Oak Ridge Affiliated Universities [1980] Industrial Energy Use Data Book, Garland STPM Press, New York.

Oi, Walter Y. [1961] "The Desirability of Price Instability Under Perfect Competition", Econometrica, Vol. 29, pp 58-64.

Oi, Walter Y. [1972] "The Consumer Does Benefit from Feasible Price Stability: A Comment", The Quarterly Journal of Economics, p 494.

Panzar, John C. [1976] "A Neoclassical Approach to Peak Load Pricing", The Bell Journal of Economics, Vol. 7, pp 521-530.

Panzar, John C. and Robert D. Willig [1979] "Theoretical determinants of the industrial demand for electricity by time-of-day", Journal of Econometrics, Vol. 9, Nos. 1/2, January 1979. (Also in Electric Power Research Institute report EA-578-SR [with comments]).

Pickel, Frederick H. [1982] "Cogeneration and Utility Planning", Ph.D. Dissertation, Massachusetts Institute of Technology, June 1982.

RAND [1979] "Real Time Pricing and the Future of the Electricity Industry", Unpublished memo (by Derek McKay?), October 1979.

Reynolds, S. P. and T. E. Creighton, Jr. [1980] "Time-of-Use Rates for Very Large Customers on the Pacific Gas and Electric Company System", IEEE Transactions on Power Apparatus and Systems, Vol. PAS-99, No. 1, Jan/Feb. 1980, pp 147-151.

Rogerson, William P. [1980] "Aggregate Expected Consumer Surplus as a Welfare Index with an Application to Price Stabilization", Econometrica, Vol. 48, pp 423-436.

Rotemberg, Julio J. [1981] "Monopolistic Price Adjustment and Aggregate Output", MIT Sloan School Working Paper WP 1186-81.

Samuelson, Paul A. [1972] "The Consumer Does Benefit from Feasible Price Stability", The Quarterly Journal of Economics, p 476.

Scherer, Charles R. [1976] "Estimating Peak and Off-Peak Marginal Costs for an Electric Power System: An Ex-Ante Approach", The Bell Journal of Economics and Management Science, Vol 7, Autumn 1976, pp 575-601.

Scherer, Charles R. [1977] Estimating Electric Power System Marginal Costs, North Holland Publishing Co, Amsterdam.

Schmalensee, Richard [1979] The Control of Natural Monopolies, Lexington Books, Lexington, Mass.

Schuler, Richard E. and Benjamin F. Hobbs [1981] "The Consequences of Alternative Organizations of the Electric Utility Industry", Presented at 1981 Annual Meeting, American Economic Association.

Schwepe, F. [1978] "Power Systems 2000", IEEE Spectrum, July 1978.

Schwepe, F. et al [1980] "Homeostatic Utility Control", IEEE Transactions on Power Apparatus and Systems, Vol. PAS-99, No. 3, May/June 1980.

- Sheshinski, Eytan and Yoram Weiss [1977] "Inflation and Costs of Price Adjustment", Review of Economic Studies, pp 287-303.
- Steiner, Peter O. [1957] "Peak Loads and Efficient Pricing" Quarterly Journal of Economics, Vol. 71, pp 585-610.
- Stigler, George J. and James K. Kindahl [1970] The Behavior of Industrial Prices, National Bureau of Economic Research, New York, HB 235.06. S855.
- Sturgeson, James I. [1980] "Discussion Series on PURPA Related Topics", U. S. Department of Energy, DOE/RG/00335.
- Takayama, Takashi and George Judge [1971] Spatial and Temporal Price and Allocation Models, North Holland Publishing Company, Amsterdam.
- Tschirhart, J. and F. Jen [1979] "Behavior of a Monopoly Offering Interruptible Service", Bell Journal of Economics, Vol. 10, pp 244-258.
- Tsiksiklis, John N. [1979] "Linear Optimization Problems with Dynamical Structure", Senior Thesis, Massachusetts Institute of Technology, December 1979.
- Turvey, Ralph and Dennis Anderson [1977] Electricity Economics, Johns Hopkins University Press, Baltimore.
- Vardi, Joseph and Benjamin Avi-Itzhak [1981] Electric Energy Generation: Economics Reliability and Rates, The MIT Press, Cambridge.
- Vickrey, William [1955] "Some Implications of Marginal Cost Pricing for Public Utilities", Papers and Proceedings of the American Economic Review, Vol. 45, pp 605-620.
- Vickrey, William [1971] "Responsive Pricing of Public Utility Services", Bell Journal of Economics and Management Science, Vol. 2, No. 1, pp 337-346.
- Vickrey, William [1978] "Efficient Pricing Under Regulation: The Case of Responsive Pricing as a Substitute for Interruptible Power Contracts", unpublished manuscript, Columbia University.
- Wachter, M. and Oliver Williamson [1978] "Obligational Markets and the Mechanics of Inflation", The Bell Journal of Economics, vol. 9, pp 549-571.
- Weitzman, Martin L. [1974] "Prices vs. Quantities", Review of Economic Studies, October 1974.

Wenders, J. T. [1976] "Peak Load Pricing in the Electricity Industry", The Bell Journal of Economics, Vol. 7, pp 232-241.

Williamson, Oliver E. [1979] "Transaction-Cost Economics: The Governance of Contractual Relations", Journal of Law and Economics, pp 233-261.