Demand Trading: Building Liquidity

Technical Report

Demand Trading: Building Liquidity

1001635

Final Report, December 2002

EPRI Project Manager
W. M. Smith
P. C. Meagher

EPRI • 3412 Hillview Avenue, Palo Alto, California 94304 • PO Box 10412, Palo Alto, California 94303 • USA
800.313.3774 • 650.855.2121 • askepri@epri.com • www.epri.com

DISCLAIMER OF WARRANTIES AND LIMITATION OF LIABILITIES

ORGANIZATION(S) THAT PREPARED THIS DOCUMENT

Apogee Interactive, Inc.

ORDERING INFORMATION

Requests for copies of this report should be directed to EPRI Orders and Conferences, 1355 Willow Way, Suite 278, Concord, CA 94520, (800) 313-3774, press 2 or internally x5379, (925) 609-9169, (925) 609-1310 (fax).

CITATIONS

This report was prepared by

Apogee Interactive, Inc.
2100 East Exchange Place
Suite 100
Tucker, GA 30084

Principal Investigator
J. Gilbert

This report describes research sponsored by EPRI.

The report is a corporate document that should be cited in the literature in the following manner:

Demand Trading: Building Liquidity, EPRI, Palo Alto, CA: 2002. 1001635.

REPORT SUMMARY

Demand trading holds substantial promise as a mechanism for efficiently integrating demand-response resources into regional power markets. However, regulatory uncertainty, the lack of proper price signals, limited progress toward standardization, problems in supply-side markets, and other factors have produced illiquidity in demand-trading markets and stalled the expansion of demand-response resources. This report shows how key obstacles to demand trading can be overcome, including how to remove the uncertainty associated with returns on demand-response investments, thus opening the way toward greater market stability.

Background

The Federal Energy Regulatory Commission's draft Standard Market Design has formally recognized the stabilizing benefits of incorporating demand-response resources more fully into regional power markets. These resources can lead to a better supply-demand balance and mitigate the types of price spikes and power shortages that have affected various regions of the United States in recent years. The path toward achieving these goals, however, is far from clear. In fact, the demand-response resource base is shrinking in many parts of the United States despite the real risks of generation and transmission constraints in regional power markets within the next several years.

Objectives

To investigate the causes for the current lack of liquidity in demand-trading markets; to develop market-oriented approaches for building liquidity into demand trading and for expanding the role of demand-response resources in hedging against future market risks.

Approach

The project team combined its expertise in demand response, discussions with industry experts, and analysis of recent power market experiences to analyze both the problems and the potential for demand trading. The effort developed innovative concepts that can enhance liquidity and improve the effectiveness of demand-trading markets.

Results

This report considers new analysis approaches and market mechanisms that can assist in expanding demand trading's role during the current incomplete transition toward competitive power markets in the United States. It considers five factors that will be needed to bring greater liquidity to demand-trading markets:

Standardized definitions for demand trading: Standardization is a basic requirement for liquidity in any market because buyers and sellers must have a common understanding of what is being transacted. This document discusses ways to formalize the evaluation of customer baselines and

demand-response capabilities into commodity-like attributes that can be sold, resold, aggregated, and settled in open, competitive markets.

Valuing demand response and trading across traditional boundaries: The fact that different types of organizations in a region will place different values on demand-response resources will help to drive demand trading in the future. To be most effective, this trading should be in open markets that allow access to multiple parties, not just the load-serving entity or the independent system operator/regional transmission organization that defines the traditional boundary for demand-response activities.

Appropriate analysis approaches for demand trading: Given the current market environment and the intermittent nature of demand-response benefits, demand-trading investments can be hard to justify using traditional economic approaches. A complete analysis of demand trading should include its value as a hedge against price volatility and system constraints. This report borrows the concept of "air worthiness" from the aircraft industry and applies it to the power industry to develop new metrics for the evaluation of demand-trading investments.

Approaches for dealing with risk: Unhedged price risk in the power industry presents a danger not only to energy companies that can be caught in an economic squeeze between retail and wholesale markets, but also to regional economies and all market participants. This report considers approaches for managing this type of risk, including insurance requirements and risk-pooling mechanisms.

Mechanisms for developing forward markets: The energy industry traditionally relies on forward markets for price signals, but these markets are not currently functioning well. Without the price signals needed to encourage demand trading, alternative mechanisms can be considered for "making the market forward." One possible approach would be the development of regional demand-response reserve banks. These would be places where customers could deposit their demand-response capabilities in exchange for periodic interest payments (reservation fees) plus use-transaction fees as the resources are used.

EPRI Perspective

To build liquidity for demand trading, new approaches will be needed to help demand trading achieve its full potential during the transition toward more-competitive power markets. This document seeks to spark action toward the goal of reaping the benefits of demand trading for regional power markets.

Keywords

Demand trading
Demand response
Price response
Risk management

EXECUTIVE SUMMARY

This document, ***Demand Trading: Building Liquidity***, begins where EPRI's ***Demand Trading Toolkit*** (EPRI 1006017) left off. That document explained the theory and practice of demand trading and how it has the potential to dramatically enhance the scope of demand response in regional power markets. By more closely integrating customers' demand-reduction capabilities into these markets, demand trading can promote a better supply-demand balance and mitigate the types of price spikes and power shortages that have affected various regions of the U.S. in recent years.

This document focuses on the key issue that needs to be addressed in fully enabling demand trading's role in competitive electricity markets – building liquidity. The current lack of liquidity for demand trading, a problem exacerbated by illiquidity on the supply side of the industry, presents a difficult challenge.

The Federal Energy Regulatory Commission's draft Standard Market Design has formally recognized the stabilizing benefits of including demand response more fully into regional power markets. However, the path toward achieving this goal is far from clear. The liquidity of demand-trading markets and the future of demand response itself are intrinsically tied up in the answers to important questions such as:

- How do you define the demand-response resource?
- Who owns the resource?
- How do various entities along the electricity value chain value it?
- How can you trade it beyond the boundaries of the energy supplier?
- What mechanisms can be used to "level out" the intermittent benefits of demand response?
- What business models can lead to a greater role for demand trading?

Moreover, in areas of the U.S. where restructuring has occurred, the available demand-response resources are sometimes smaller than the amounts previously available under the jurisdiction of the vertically integrated utilities. As electricity markets opened, the hope was that free-market innovation would produce an array of demand-response approaches that could be offered along with retail electricity services. A variety of factors, however, discouraged this. Regulatory policies, the intrinsic economic risks of the retail proposition, the way traditional economic analysis approaches were applied, the high costs of customer acquisition, and other factors limited new demand-response offerings.

Currently, with most regions of the country experiencing “soft” forward markets for electricity, investments in demand-response resources can be difficult to justify using traditional economic analysis methods. The benefits are often considered as either too uncertain or too infrequent to justify building the necessary demand-response relationships and systems. As a result, many demand-side resources remain untapped and when power markets eventually go into constraint and prices rise, these resources will be unavailable for use.

Without solutions to the liquidity and other problems facing the demand side of the industry, it is likely that the U.S. will become increasingly vulnerable to the recurrence of power shortages and price spikes. What then, are the key factors needed to bring liquidity to demand trading and expand the demand-response resource? This document defines five such factors and explores how they relate to the future of demand trading:

- Standardized definitions for demand trading.
- Valuing demand response and trading across traditional boundaries.
- Appropriate analysis approaches for demand trading.
- Approaches for dealing with risk.
- Mechanisms for developing forward markets.

Standardization is a basic requirement for liquidity in any market. Buyers and sellers must be aware of what they are getting or providing, and they must be aware of their obligations and rights. Accordingly, as electric customers are approached with offers to include their demand response into retail agreements and those capabilities are traded into regional markets, it will be necessary to use standard definitions to characterize the demand-response resources. Standards must go beyond any one company’s point of view and be acceptable to all counterparties to promote commerce. Without this standardization, efforts at using demand trading will be inhibited by uncertainty and by unreasonable transaction costs. This document discusses ways to formalize the evaluation of customer baselines and demand-response capabilities into commodity-like attributes that can be sold, resold, aggregated, and settled in open, competitive markets.

Once standardized demand-response resources are defined, different market players in a region will often value these resources differently. For example, some will value a resource only as temporary relief from high prices in the short-term market, whereas others will value it primarily as firm capacity. Additionally, a market player’s perspective can change depending on whether is has a “long” or “short” forward position in the regional power market.

These differing valuations can help to drive demand-trading activity, but only if a customer’s demand-response resource is accessible through an open market. In some areas, restrictions currently prohibit such a resource from being offered to any organization other than the load-serving entity or the ISO/RTO. Eventually, however, it is expected that virtually all regions will allow demand trading across these traditional boundaries. This will provide an opportunity for those placing the greatest value on a demand-response resource to have access to it.

In general, during periods of soft forward markets, demand-side investments can be difficult to justify. One should consider, however, the inherent value of demand trading as a hedging mechanism to protect against periods of constrained power availability and high wholesale prices. While such periods may not occur frequently in any given region, their effects can still be disastrous. As history has shown, unhedged retail energy companies can be caught in an economic squeeze between high prices in the wholesale markets and low prices in retail contracts.

Demand trading presents a way for an energy company to manage its exposure to this type of price risk. By having the ability to reduce demand during critical periods, huge savings can result. This can have a big impact on the bottom line – perhaps the difference between bankruptcy and simply having a bad quarter.

In evaluating demand trading's value as a hedge, one can consider the concept of "air worthiness." In the aircraft industry, this is defined as the ability to anticipate threatening situations and withstand unavoidable turbulence without destroying the airframe or harming those inside. By considering this concept in the context of the power industry, this report develops the metric of "survivability," which defines an energy company's ability to survive and gracefully exit the worst-case scenario in any one year. Another metric discussed is "staying power," which considers the cash requirement to remain solvent through a multi-year run of bad luck.

These metrics focus on the energy companies that supply retail power, but it is important to remember that the risk associated with price spikes and power shortages extend beyond the energy industry. As much of the country has experienced, it can affect the entire population and the economy within a region. Given this, the "insurance" provided by demand trading has broad social benefits.

In a perfect market, the implementation of this type of insurance could result in higher prices for customers that are inflexible in their energy use, compared to those that are flexible (given similar energy-use patterns). Additionally, retailers that do not hedge their supply portfolios to meet their customer obligations could perhaps pay a premium when they lean on the ISO/RTO balancing markets to a high degree. In a very real sense, these retailers can jeopardize all market participants by their actions, in a similar way to driving a car without insurance.

Energy markets traditionally rely on forward markets for price signals, often using public indices such as NYMEX agreements as proxies for fair bilateral terms. Unfortunately, at this time, there are no such proxies available, and we are likely to see boom-bust cycles in the valuation of both supply-side and demand-side trades in a region. These cycles are highly disruptive because they inhibit market participants from sustaining a sufficient value proposition. Looking at the issue as if from the supply side of the equation, it is like no one wanting to build another power plant without assurance that the bills on that plant will be paid.

Given the current situation, a key question is what mechanisms can be used to develop the price signals needed to encourage demand trading. One way to move in this direction would be to eliminate the “safe haven” of non-time-varying POLR rates currently in place throughout much of the U.S. These rates make it difficult for an energy company to successfully offer a risk-differentiated rate. If POLR rates were required to be real-time rates for larger customers and perhaps time-of-use rates for medium-sized customers, it would help to level the playing field for demand-response and demand-trading approaches. However, this approach could be difficult to implement and might not be widely accepted in the current environment.

Another mechanism that could be considered is the development of regional demand-response reserve banks. These banks would be places where customers could deposit their existing demand-response capabilities in exchange for periodic interest payments (the reservation fees) plus use-transaction fees as these resources are used. These same banks would also lend money for investments in additional demand-response resources, similar to the way homes and business investments are financed. Figure ES-1 visually portrays how such a system could work.

New approaches are needed to help demand-trading markets function with greater liquidity during the transition toward more-competitive power markets. Some of the steps along this transition will be challenging, but all are achievable with a concerted effort on the part of many professionals and national organizations. Hopefully this document will spark further discussion and action toward the goal of making demand trading an important part of power markets.

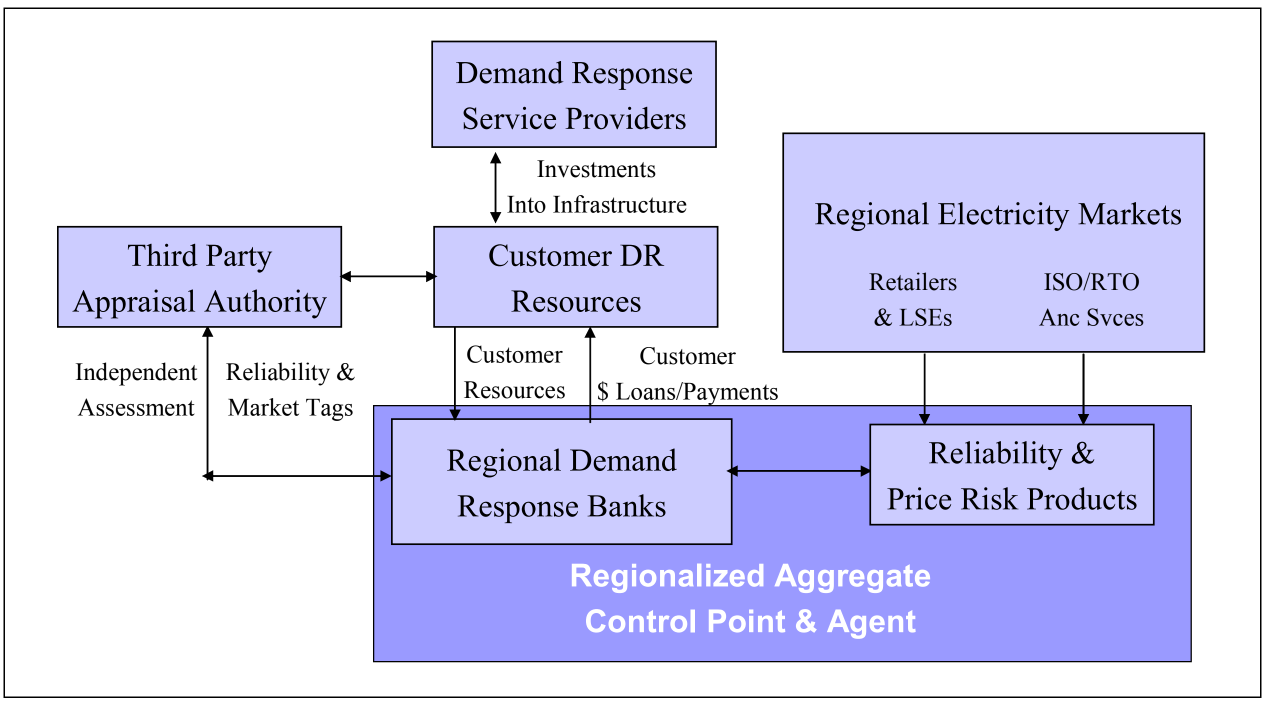

Figure ES-1
The Regional Demand-Response Reserve Bank Concept

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION AND OVERVIEW

Following a series of difficulties, including the collapse of Enron, the power industry in the U.S. is facing a serious lack of liquidity. What does that really mean? Also, given that liquidity is a key ingredient to the efficiency of an open market, how can it be increased? These are central issues covered in this document. In particular, this document will investigate how to increase liquidity in demand-trading markets.

If the reader is new to the subject of demand trading and wants to review the basics of this concept and the underlying principles that cause liquidity problems in the first place, it would be useful to review EPRI's ***Demand Trading Toolkit*** (EPRI 1006017), published in 2001. It explains the reasons energy companies use demand trading in electricity markets to either reduce purchases of higher-priced energy or sell their released power-sales obligations into regional markets.

This document moves beyond the ***Demand Trading Toolkit*** to examine some of the issues remaining to be solved to bring demand trading to maturity as an effective alternative to supply. Most importantly, it will help to codify the various mechanisms needed to create liquidity for demand-trading markets. In many ways, this is analogous to the early days of building the liquidity of currencies in the United States. The early attempts to replace barter and the currencies of gold and silver with "paper agreements" have many parallels to the problems we are encountering with the standardization of demand-trading agreements at this time[1].

Before tackling the issues surrounding the liquidity of demand-trading markets, it is instructive to look at a working definition of liquidity. Basically, it has to do with the ease with which you can exchange an item for what you want. Canadian currency is more liquid in states bordering Canada, for example, but increasingly illiquid as you move South. It simply isn't worth the trouble to try to exchange it. (Interestingly the innovation of credit cards enables the orderly transfer of funds easily between parties, virtually eliminating the need for the exchange of physical currency intermediaries.)

Why is an airline ticket so illiquid? After all, you can sell your car, boat, and household belongings on the Internet via an auction website, but you can't sell your airline ticket to

---

[1] The reader would probably enjoy and benefit from a review of this history found on the Federal Reserve Bank website http://www.frbsf.org/federalreserve/money.

someone else if you no longer need it or want to change your plans. While you can sell the ticket back to the airline and take the penalty, you don't have the right to resell that ticket to anyone else.

Why aren't there any intermediaries able to step in and create this market? The reason centers on the question of risks and responsibilities in *bilateral agreements*. The airline is simply selling you the right to occupy a seat they own, and you don't have the right to resell that seat. Various rules, risks, and rewards are intrinsically priced into the ticket and one of these rules dictates that the ticket is not transferable to another person.

Futures contracts and derivative instruments, on the other hand, are designed to be liquid – more easily traded on open markets. Gold, oil, soybeans, and pork bellies are all commodities that lend themselves to such agreements (assuming you have enough of these commodities, they are of adequate quality, and you can deliver them to accepted trading points). When commerce in such commodities is well established, buyers and sellers can use futures and derivative financial instruments to pre-arrange the transfer of physical and financial responsibilities and risks. Most trades in these commodity markets are for financial hedging (avoiding the risks of undesirable price movements). For most commodities, relatively little volume in the trading of futures actually results in the trader taking possession of the commodity.

Attempting to use these mechanisms in electricity trading has caused frustration with the current partial transition toward open, competitive markets. The recent Standard Market Design (SMD) concept from the Federal Energy Regulatory Commission (FERC) has formally recognized a need for both supply and demand to participate in regional power markets. However, succeeding at the objective of integrating customers' demand-response resources more fully into power markets is fraught with difficulties at this time.

While the supply-side professionals might lament the lack of liquidity in their markets, those working to integrate demand-response mechanisms into power markets face far less liquidity. A variety of detailed questions are being hotly debated, including:

- Who owns the demand-response resource (e.g., the customer or the serving energy company)?
- How do you define the resource (capacity, energy, firm, non-firm, …)?
- How do various entities value the resource (displaced purchases, opportunity to make money, reduced operational stresses, …)?
- When and how can the resource be traded beyond the boundaries of the energy supplier and to whom (as counterparties)?
- How can counterparties and agents be properly compensated for their roles?
- Under what conditions might a customer rightfully pledge demand response while under a curtailment or an interruptible agreement?
- What mechanisms can be used to "level out" the intermittent benefits of demand response?

- What business models can lead to a greater role for demand trading?

The liquidity of demand-trading markets and the future of demand response itself are intrinsically tied up in the answers to these questions. Some at this time would like to delay acceptable interim answers until the electricity market itself is perfected. Others would like to force fit the answer to what they can control at the moment. Hopefully, this document will help industry participants answer these questions with free-market solutions rather than force fitting demand trades into another regulated paradigm. To that end, this document suggests some new analysis approaches and details some options for expanding the role of demand trading in rapidly evolving power markets.

The remainder of this chapter summarizes the current situation of demand trading and then goes on to outline the key factors needed to promote demand-trading liquidity.

The Current Situation

There is widespread agreement that demand response is an important strategic asset that can mitigate price and volume risks in today's energy business. To demonstrate how demand trading can provide these benefits, let's begin with a simple example.

Assume that a customer has signed a contract for power supply for just one year of service at some fixed price per kWh. Facing competitive pressure and wishing to keep its prices as low as possible, the supplier has not paid a premium for a hedging strategy that would guard against wholesale price spikes. The supplier therefore faces both price and volume risk in this agreement given that it has no assurance about how much the customer will buy nor what the acquisition costs will be for the power consumed.

In the best of cases, when the weather is mild and the energy markets are "soft," there may be ample headroom[2]. However, the competitive pressures noted earlier may take most of the margin out of this situation and the market players may achieve only modest rewards. For the average situation, when the weather is normal and the markets are working properly, the retailer can squeak by with some small level of profitability, or perhaps an acceptable loss that will hopefully be made up in ancillary products and services offered to these customers as time goes on.

But, when things turn for the worse, two compounding factors devastate the economics. The volume (kW and kWh used) typically goes up (since the weather is hotter or colder than normal) and probably exposes the retailer to volume risk at the same time that acquisition prices are

[2] The term headroom is generally used to describe the difference in price between the price the customer pays the retailer and the price at which the supply is acquired. Headroom is a necessary but far from sufficient criteria for retail success. Other costs include administrative costs, including the acquisition of the customer, metering, billing, and customer care. Headroom itself may not be available when regional wholesale prices reflect constraints and the only successful retail proposition may be the use of demand-response trades. At this time, many retailers are simply closing their doors rather than attempting to solve the problem with demand trading. The reasons for this are among the key topics discussed in this document.

probably much higher than expected. The retailer is losing a lot on each kWh and this is compounded by higher volumes. As a result, it is likely that just one season of this "abnormally worse weather" will produce so much red ink that future years of normal weather simply cannot make it up over time. Given that the only alternative to facing this loss is to "slam" a customer back to the provider of last resort (POLR) at traditional cost-of-service ratemaking, the retailer loses both the relationship and the possibility of expanding the relationship in the future.

The innovative retailer might have agreed to offer the customer a demand-response approach with the promise that the economic benefit of any demand reductions during periods of high price would be shared with the customer. For example, rather than selling the customer electricity at $0.06 per kWh and buying it in the regional market at $0.20 or more per kWh during period of price spikes, the customer might receive a $0.10 per kWh credit for demand reductions on its bill. Since the price spikes in a region are often at even higher values than the one noted here, customers could possibly achieve a rather low annual equivalent price by participating in demand-response events for less than 100 hours a year. In fact, energy companies who have exercised this model have been able to offer customers 15-30% discounts off standard tariffs if those customers can drop off the grid completely for this number of hours a year.

Unfortunately, free market energy retailers have had enormous difficulties acquiring customers due to the lack of headroom and heavy competition for the most desirable customers. The result has been a pattern of cutting the price to the bone to get volume, hoping to add value at some time in the future.

The regional independent system operators (ISOs) have also recognized that demand-response resources can assist in reserve and reliability markets. In addition, since these ISOs operate balancing markets for market participants to buy and sell to meet their actual vs. planned schedules, they have seen that demand-response impacts can influence the final cleared price in the Dutch auctions for those markets, potentially saving much more than the displaced transaction itself[3]. This mitigation of price spikes presents a substantial value to society.

One can easily argue that there are other societal benefits of demand response as well. For example, due to environmental benefits, the load impacts of customers willing to turn off lights are superior to the supply-side equivalent. In addition, the reduction in peak loads has potential benefits on system stresses (transmission line, distribution, and transformer overheating and consequent failure) and can sometimes delay or avoid the need to upgrade T&D assets.

Demand trading has the potential to greatly expand the application of demand-response alternatives in regional electricity markets. As its name implies, trading involves the buying and

[3] This effect has been documented by many professionals and is also highlighted in the author's paper to the FERC on the need for regulatory changes – Customer Demand Response (The Four Not so Easy P's) – presented at the FERC/DOE Workshop on Demand Response in Washington, D.C. on February 14, 2002.

selling of a resource. Trading a customer's ability to change load shape places the customer in the position of seller and the energy retailer as the first potential buyer. If that retailer sees no need to purchase this capability for its own use, that retailer would have to find a counterparty to the transaction. However, when one attempts to implement a trade beyond the energy company's own scheduled vs. resource balance parameters, numerous questions and concerns can cloud the transaction and bring with it significant financial risk. Very simply, a customer's choice to reduce loads is not the same as a generator's output. Metering alone fails to answer the question since it is possible (and likely in some cases) that the customer has shifted use to other periods of time and thereby raised its energy use during those periods. Standardization of a customer's demand response effects – the net load shape impacts – is required to communicate that capability beyond the energy company's borders.

In addition, aggregation improves the confidence one has in the net load shape effect since some customers will perform better than pledged while others will fail to live up to their pledged demand reductions. Some customers can respond automatically to price signals and could literally be put on automated generation control equivalents at the ISO level and receive the same payment as generators. However, most customers want some level of voluntary interaction – they generally want to consider the price and decide whether to take actions.

Of course, the intermediaries along the way also need their fair share of the opportunity as well. And all of this has to be conducted in a way that ensures a reasonable balance between benefits and risks for each member of the value chain. Unfortunately, this chain is rather disjointed and broken in most areas of the country at the moment.

Customers have historically looked to their regulated energy supplier for these types of programs and were often pre-paid for their capabilities through some form of a demand-charge reduction. While these types of interruptible/curtailment agreements were in place and not exercised, the customers often saw them simply as a discount. When they began to be exercised, some of these same customers complained and attempted to get off the rate itself. For example, when energy companies used these programs to avert high-priced wholesale power purchases during the heat waves of 1998 and 1999, some of the participating customers literally sued to get a larger slice of the pie.

The Difficulties Resulting from an Incomplete Market Transition

As electricity markets moved toward open competition, the hope was that reduced levels of regulation would spark free-market innovation and that retailers would bundle demand-response approaches into their retail propositions. Some retailers attempted to do so, but regulatory policies, the intrinsic economic risks of the retail proposition, the way traditional economic analysis approaches were applied, high costs of customer acquisition, customers' desires for simplicity, and other factors subverted their efforts. If there were a robust retail market, greater free-market innovation probably would have commenced. However, given the constraints in the retail market, relatively few creative demand-response products and services were offered.

One of the problems faced by retailers in today's incomplete transition to retail competition is POLR pricing which often allows customers a "safe place to hide." In most areas of the United States, there are still at least several years of these price protections left. As a result, customers continue to be offered POLR prices that are set on traditional cost-recovery rate setting models which do not adequately reflect risk premiums associated with a competitive market. This can make it difficult for a retailer to compete. Matching the POLR price can leave the retailer in a very risky position. Alternatively, building the prudent risk premium into prices can leave the retailer with an uncompetitive offering.

Making matters worse, when the forward markets are "soft" as they are in the current environment, investments in demand-response resources can be difficult to justify economically. As a result, when energy markets eventually go into constraint and prices rise, energy retailers are likely to find themselves in an economic squeeze. Very simply put, today's energy price forecasts make energy retailers feel they can not afford to prepare for the bad times ahead because the economic benefits are either too uncertain or infrequent to justify building a demand-response relationship with customers.

Compounding the situation is the attitude of many energy customers, who tend to have very short memories. They will accept real-time pricing when prices are soft and when doing so yields a lower price than the bundled tariff, but then jump back (if allowed) to the protection of the POLR tariff when prices spike. Similarly, customers generally prefer interruptible and curtailment agreements when they are not exercised often, and on the contrary prefer voluntary agreements when they are exercised often. Customers are prone to jump between agreements (if they are permitted to do so) based upon which gives them the biggest bang for the buck at the time.

As noted earlier, demand trading is also being thwarted by the lack of standards in demand response. Without standards, the customer's ability to change load shape becomes illiquid – when the retailer attempts to communicate the terms upon which they will entertain a demand-response trade, the uncertainty in economic valuation causes substantial confusion.

The result is predictable. Demand-trading capabilities all across the U.S. are declining at this time. While significant amounts of demand response are available at the ISO level in certain areas, it is instructive to recognize that these amounts are sometimes a small fraction of what was available under the vertically integrated utility's jurisdiction. For example, the 4,000 MW of demand-trading resource in ERCOT before the market opened has now dropped to less than 500 MW.

Ironically, many of the best examples of creative demand trading are coming from areas in the country where retail choice is still a bit off in time. The largest volumes of creative demand trades have come in the Pacific Northwest and the Southeast where many new approaches are being considered. The Bonneville Power Administration is using demand trading for both energy market and transmission load-relief purposes. Entergy permits Louisiana customers to offer demand reductions at a price, and can actually counter those to settle on a cleared price. Georgia Power can offer specific customers at points of congestion a price and monitor their aggregate performance within that zone.

If demand-trading liquidity problems are not solved, it is possible that the continued push towards restructuring will continue to erode demand-trading resources and bring us right back into the reliability crises various parts of the country faced in 1999 and 2000.

What Is Needed to Make Demand Trading Liquid?

This section touches on the key factors needed to bring liquidity to demand trading and introduces the topics that will be covered in more detail in later chapters. The factors are:

- Standardized definitions for demand trading.
- Valuing demand response and trading across traditional boundaries.
- Appropriate analysis approaches for demand trading.
- Approaches for dealing with risk.
- Mechanisms for developing forward markets.

Standardized Definitions for Demand Trading

Demand trading is similar to the process of building a house. The buyer asks a builder to provide a quote. The builder is apprehensive and wants the job, but knows that there are too many attributes that need definition before a final quote can be made. Plus, the builder isn't sure the customer understands the tradeoffs between quality and value. Given that the builder is probably going to receive a percentage of the total cost as compensation, the builder may want to "get the customer committed" with a low (but realistic) ballpark estimate based upon examples of what the builder has built in the past. This pushes the question about what the real costs will be to a later time, presuming the customer doesn't want what the builder has built in the past.

Then, when the customer is committed, the builder will build exactly what the customer wants, with all the extra features requested. Eventually the customer may find that when selling the house, the new buyer doesn't value the extra features they incorporated into the house. So, what is a house worth? How is this different, better, or worse than the speculative builder who understands the longer-term valuation process? And, why would builders speculate anyway rather than simply wait around for the custom orders?

Anyone who has experience with the free market will attest that the proof of valuation in a market is whatever someone is willing to pay for it, not just once, but over and over again.

This description is actually pretty close to the way demand response is being acquired and traded today. All market counterparties are thinking something a bit different and there is no standard around which to determine "comparables" for valuation. A buyer decides what they want, contacts the customers they can influence, signs up the resource they can, and the resource is probably traded infrequently, if at all. Then, the customer is abandoned, and the next builder attempts to build something with their resources. In many cases, the customer resources are not able to achieve their full economic value. Part of this illiquidity is due to the lack of standards in the design of the demand-response platforms, and part of it has to do with the buyer wanting the resource to look like generation.

We also need to be more precise in our language about demand response. It is common to hear discussions about real-time pricing and demand bidding and yet there are no standards to the meanings of these terms. For example:

- Where does the "price" come from in real-time pricing (calculated or some market proxy)? Does a zonal clearing price really correspond to a customer's price?
- How can that price be correct given the energy company may not know how much the customer is going to buy? After all, if all the customers responded to the same high price, that price would be wrong just because the customers responded to it.
- Does demand bidding assume the customer is a price taker or can and should customers bid in their willingness to curtail and be considered in the same security-constrained dispatch model as the supply side?

Presently, the definitions of customer baselines and other demand-response parameters are left to somewhat arbitrary and potentially contentious sets of rules. Part of the goal of this document is to help formalize the evaluation of customer baselines and demand-response capabilities into commodity-like attributes that can be sold, resold, aggregated, and settled in open, competitive markets. The aggregated resale and clearing across multiple market participants takes multilateral confidence in the demand-response capabilities as a currency.

Chapter 2 will consider the underlying attributes of demand reductions and describe these reductions so that trading instruments and structured transactions can be crafted properly.

Valuing Demand Response and Trading Across Traditional Boundaries

What is the value of peak-load reductions in any given year? This can depend on the perspective of who is determining the value. Some are interested in "firm resources" that are the equivalent of installed generation capacity. Others are simply interested in temporary relief from what they feel are high prices in the short-term energy markets. These are two very different perspectives.

Nearly everyone today seems to believe that demand response is an essential element to making electricity markets work efficiently. As partial justification for this perspective, they note that wholesale electricity markets settle at "unreasonably" high clearing prices when demand is inelastic. However, key questions exist about the value that customers should be paid, and about paying more or less for given types of demand-response resources.

Capacity value is derived from avoiding the costs of building additional generating capacity. Clearly the key question to market counterparties is whether demand response can avoid the need to build capacity. When it comes to emergency reserves, certain options seem reasonable as alternatives (e.g., controllable water heaters, etc.) while in other instances, the only comparable customer demand-response mechanism would be with the operation of a back-up generator on its site.

Energy value focuses on the arbitrage between what the customer might normally pay for energy and the wholesale-displaced price. While a few customers might participate in this market routinely in a real-time pricing program, most will not.

Experiences gained in operating demand-response bidding programs around the United States for the past few years indicate there is quite a bit of valuation variation in every dimension of this resource. One could argue that it should simply go to the highest bidder and that customers should link directly to counterparties. However, such arguments fail to recognize that most customers need an agent relationship and a simplified market interface to participate as alternatives to supply.

Another aspect of the way energy retailers consummate and value trades with customers involves the position they have in the market. They most often accept or reject aggregated capabilities, deciding whether the pledged aggregate in response to price can beneficially meet their particular forward energy needs.

This evaluation is based upon the "book" of transactions (collection of agreements) along with the energy company's forecasted needs. One should not assume the energy company has the view of an arbitrageur where all price and volume efficiency improvement opportunities are worth taking. When energy retailers "go long" and buy large blocks of power for all their planned needs out into the future at seemingly bargain basement prices (as they are prone to do at the moment), they often become lethargic about developing demand-trading options.

Similarly, there are also retailers who will be tempted to gamble that prices will be even lower on average in the spot markets and therefore might avoid forward agreements altogether. This is the equivalent of being "naked" in both volume and price, and can be devastating to a company if they find they "placed the wrong bet."

Unfortunately, the energy retailing game is quite unlike many games of chance played in Las Vegas where only a small edge is given to the house. Retailers are betting the company when they have naked positions. However, if the bet pays off, they can look like heroes. Wall Street does not yet understand how to evaluate the risks of this business and how professionals can best manage them.

As noted earlier, another issue important to demand trading is whether a customer wishing to offer demand reductions can offer them to organizations other than the load-serving entity (LSE) or the ISO/RTO. In many situations, restrictions currently prohibit such activity. In the longer term, however, allowing access by various market players in a region to a customer's demand-response resources will allow for increased efficiency. This trading across the traditional boundaries will provide an opportunity for those placing the greatest value on a demand-response resource to have access to it.

The topics of valuing demand response and trading across traditional boundaries will be covered in greater detail in Chapter 3.

Appropriate Analysis Approaches for Demand Trading

A proud son graduates from a college of agriculture and returns home to work with his dad on the family farm. After being back just a few days he gives his dad a list of 10 things that would save money and improve productivity. His dad thanks him and summarizes the situation by saying: "Son, I already know 20 ways I can do that. Haven't gotten around to those either, and don't have any plan to."

Much to the chagrin of theoretical economists, business perspectives are often driven by fear, pain, and greed – not simply costs and benefits. In most cases, things have to get pretty far off center to move people from apathy to action.

The highest levels of project activity occur when fear, pain, and greed are all working at the same time. We certainly have all seen that occur in energy markets in recent years. But when all three factors are absent, relatively little happens. The "opportunity" (or error) signal that precipitates a decision to act comes from much more than the economics of the situation. It is a synthesis of two factors – the perception of the opportunity or problem and the tolerance the person has to that perceived error signal. Many mature and efficient markets capture even the smallest of these error signals. Inefficient and immature markets can fail to capture even large error signals. These error signals are similar to the arbitrage opportunity in the electricity value chain.

The way demand-trade economics are portrayed has a lot to do with the perception of the opportunity itself. For example, when developing "business cases" for this resource, counterparties will generally have different perceptions of risks, costs of capital, abilities to capture opportunities, and other factors. As a result, while it is tempting to generalize about economic perspectives of various market players, doing so can be misleading.

For example, saving money sounds like a solid motivator, but potentially doesn't matter for those in a cost-recovery mindset. In addition, saving money that transfers funds into other departmental budgets can actually be frustrating (such as one school installing energy-efficient devices and having the savings flow to the whole school system because the school does not pay its own energy bills). We always have to consider the incentives and not just opportunities.

Therefore, as we develop the economic metrics for the evaluation of demand trading, be mindful that incentives and disincentives for action can matter more than the return on investment. This will help market counterparties understand the natural bias in economic evaluations and hopefully help all do a better job of unemotionally expressing risks and rewards in customer demand-trading opportunities.

Chapter 4 will investigate some of the traditional economic analysis approaches and consider their applicability to analyzing demand trading. This chapter then suggests a model for use in evaluating demand trading that follows the Civil Aeronautics Board's model of "air worthiness." The airline industry is an interesting analogy in that airlines are not permitted to carry passengers just because they have approved and inspected aircraft. They have to prove they are financially sound and have business procedures and personnel to support their flight schedules.

Chapter 4 also illustrates several new ways that energy companies and investors can evaluate hedging using demand-trading capabilities. One new metric for this type of analysis is a "survivability" which considers whether an energy company can survive and gracefully exit the worst-case scenario in any one year. Another metric discussed is "staying power," which considers the cash requirement to remain solvent through a multi-year run of bad luck.

The end result of these new ways to look at demand trading illustrates that while the demand-response resource may be small in relation to the total forward energy needs, it can be extremely significant in mitigating many price and volume risks. One can imagine a day when the only way an energy retailer can afford to compete is with a portfolio of customers offering demand-trading capabilities. It is possible that some retailers might buy this resource forward at a low price and create a first-mover advantage. Such activity, however, is less likely in light of the current problems in the capital markets for retailers.

Approaches for Dealing with Risk

Price risk plays a central role in restructured power markets and in the field of demand trading. Competitive market participants try to identify, quantify, and assign price risks. In the old regulatory model, the risk an individual electricity customer placed on the grid was never adequately identified because the bundling and averaging of prices didn't need to reflect it. In more-competitive power markets, the lack of clarity on price risk can become extremely costly in situations where reserve margins and wires constraints bring about high prices.

Competitive markets generally assign a price risk to situations that prove to be more costly to serve. Customer groups or segments, once identified and quantified, tend to be segregated and the price-risk costs are allocated appropriately. For example, there are some common situations where people who place higher risks on retailers pay higher prices. Here are a few:

- Health care and life insurance for smokers,
- Auto insurance for teenagers and others who tend to drive carelessly, or
- Changing your mind at the last minute in air travel.

Risk naturally migrates and disaggregates in truly competitive markets, and that eventually results in price-risk differentiated products and services. This is a good thing in most markets because it encourages innovation and efficiency in the markets themselves. However, at this time, faulty market designs, regulatory protections, and other factors have temporarily slowed this process in the power industry.

In a perfect market, customers who are inflexible in their energy use should pay more than those who are flexible (given similar energy-use patterns). Retailers who do not hedge their supply portfolios to meet their customer obligations should perhaps pay a premium when they lean on the ISO/RTO balancing markets to a high degree. In a very real sense, these retailers can jeopardize all market participants by their actions, in a similar way to driving a car without insurance.

Today's retailers have enough trouble just getting customers to switch away from their regulated energy providers. When they do, and when wholesale price spikes cause an economic squeeze, the easy solution is to dump them back to the POLR. We need a better interim structure to offer retailers an attractive way to deal with risks.

A possible structure is through risk pooling. Risk pools are common today in several lines of insurance. The most common are healthcare and automobile coverage. Interestingly, healthcare coverage is usually optional but automobile coverage is mandatory. The reason may be obvious, but deserves mention. Individuals simply take on their own healthcare risk, but they expose others to risks if they do not carry automobile coverage. This is an extremely important point in demand-trading capability. Since a lack of this resource exposes all market participants to price risks, one way to reduce risk would be to require "proof of insurance" as a market requirement.

Chapter 5 of this document offers specific guidance to those trying to be creative regarding the topic of managing risk.

Mechanisms for Developing Forward Markets

Capital markets are absolutely amazing to watch. Free markets raise significant capital and put it at risk if forward market signals indicate a reasonable return on investment. Unfortunately, the reverse is true as well. If the forward markets are either uncertain or fail to suggest an adequate return on investment, risk capital can be hard to find.

Energy markets traditionally rely on forward markets for price signals, often using public indices such as NYMEX agreements as proxies for fair bilateral terms. Unfortunately, at this time, there are no public forward price signal proxies for power and the mood in the market is far from optimistic. In addition, even the mood about capacity to meet day-ahead, hour-ahead, and reliability requirements seems soft. Forward markets need to be rebuilt to correct the current situation.

Without forward markets, we are bound to see boom-bust cycles in the valuation of both supply-side and demand-side trades in a region. These cycles are highly disruptive and counterproductive to demand trading because they inhibit market participants from sustaining a sufficient value proposition. Customers will not tolerate offers that produce significant benefits only occasionally. Therefore, left to the vagaries of the market and the weather, demand trading will suffer until a forward market is developed. Somehow long-run avoided costs must be set on a broad enough scale, with a long enough viewpoint, that value assurance for demand response is reasonable. Said another way, looking at the issue as if from the supply side of the equation, it is like no one wanting to build another power plant without assurance that the bills on that plant will be paid.

Once the price signals are clear for seasonal, day-ahead, and hour-ahead demand-response capabilities, the resources can be developed and brought to market. Clever and inventive professionals can certainly imagine a robust range of structured transactions between themselves and their customers that might include every type of relationship in this potential market: capacity, energy, and ancillary services.

Price-based solutions for moving toward forward markets would be the preferred approaches. One way to move in this direction would be to eliminate the "safe haven" of non-time-varying POLR rates currently in place throughout much of the U.S. As noted earlier, these rates make it very difficult for a retailer to successfully offer a risk-differentiated rate. If POLR rates were required to be real-time rates for larger customers and perhaps time-of-use rates for medium-sized customers, it would help to level the playing field for demand-response and demand-trading approaches. However, this approach could be difficult to implement and might not be widely accepted in the current environment.

If it is not likely that price-based solutions can be implemented for markets in the near term, then other (reliability-based) mechanisms can be considered. One such mechanism is the development of regional demand-response reserve banks. These banks would be places where customers can deposit their existing demand-response capabilities in exchange for periodic interest payments (the reservation fees) plus use-transaction fees as these resources are used. These same banks would also be places where customers can borrow money to invest in additional demand-response resources, similar to the way customers finance their homes and business investments. Chapter 6 covers these and other topics related to forward markets in greater detail.

Overall Goal

This document helps to formalize and define demand-trading fundamentals and mechanisms. It defines the demand response that customers can offer and suggests how it can be converted to a tradable resource through standardization and market-clearing mechanisms. In this way, an LSE or curtailment service provider serving customers with demand trades can bring that capability to the market or to any outsider who values that resource. This document shows how information coordination and financial equity can help to protect all parties to these transactions.

This document also examines the electricity value chain that includes customers, energy providers, traders, and other market players. It looks at options that give customers and others along the value chain assurance of reasonable annual benefits from demand response. Given the significant public benefits available from demand trading, the costs of not encouraging this resource and simply watching the boom-bust cycle replace the old least-cost planning and integrated resource planning models is far from desirable.

This document discusses a wide variety of topics related to creating liquidity. It investigates the human dimensions of this issue, looks closely at what makes different people in the market respond to prices, and considers how aggregation and coordination can make it easier for them to assist in linking supply and demand.

This document focuses on workable solutions with a long-term eye on what will clear a path around the debris of an incomplete transition to an open, competitive market for electricity. For the nearer term, it suggests approaches that can assist in the development of demand-trading markets during the transition.

It seems clear that many of today's electricity market problems can be solved with appropriate price signals and protections against improper conduct. Demand-trading approaches can be important parts of the solution as well, and these approaches can be encouraged by stabilizing the valuation for demand response while also permitting the market to speak to any and all customers.

How can these goals be accomplished? By clearly defining what is being traded, crafting structured transactions that can be executed as alternatives to supply, portraying the economics of acting on these agreements to all parties in the market, and establishing risk pooling mechanisms.

2 STANDARDIZED DEFINITIONS FOR DEMAND TRADING

What is 1000 cubic feet of natural gas worth and can you use it if you buy it? Standardization in natural gas markets makes free trade of that commodity possible. Buyers and sellers are aware of their obligations and rights. They know what they are getting. Similarly, as electric customers are approached with offers to include their demand response into retail agreements and those capabilities are traded into regional markets, it will be necessary to use standard definitions to characterize the demand-response resources. Standards must go beyond any one retailer's point of view and become acceptable to all counterparties for them to promote commerce. Efforts at using demand response can therefore be thwarted by a lack of standardization in language, communication of load shape and price risk objectives, as well as evaluation metrics along and across the entire electricity value chain in a region. It is therefore an imperative to create definitions and formalisms for demand trading that can communicate capabilities and risks to all counterparties in open, competitive markets.

The supply side has well-established mechanisms to characterize what is generated. Meter data are conclusive and in many cases, available on demand. In many jurisdictions it can be confirmed every few seconds. Generator output, while variable, can be forecast with relatively high accuracy (assuming it operates properly), and any deviations from contractual levels can be dealt with precisely. The same is not true with most demand-trading mechanisms. How do you measure something that didn't happen versus something that did, in a way that can be unilaterally understood and traded?

The reasons for this difficulty are not surprising. Customer load shapes are most often based on a combination of comfort decisions, operational factors, occupancy (or some other activity attribute), weather conditions, etc. Some of these are random (such as the starting and stopping of the refrigerator or air conditioner in any given hour of the day), while others are clearly the result of use decisions (such as turning lights off or on).

The central concept of trading demand reductions is to influence the controllable elements of a load shape to produce some net economic benefit. However, anecdotal words about a customer's baseline and its demand-response capability as being "repeatable" or "noisy" are inadequate to describe customer capabilities to others interested in buying such a resource. They want to know what they are getting, and to hold the market counterparties to some standard before payment.

As a result, demand response will be clearly worth less than generation if it exhibits unreliable and uncertain quantity and quality (e.g., a customer with a noisy baseline and uncertainty in the repeatability of its demand-response level). On the other hand, demand response such as control of lighting systems is likely to be more valuable than generation since it is easily measured and verified, environmentally superior than operating a fossil-fueled peaking plant, possibly already delivered in the zone, and can even be automatically linked to the market.

This chapter will consider the underlying attributes in demand response that carry the potential to increase value. It will then define how those attributes can be quantified so that a market for their value can be established. Said very simply, the chapter will characterize what the customer is selling in demand response so that a multitude of counterparties will entertain buying it.

Characterizing the Demand-Response Resource

Let's start with the simplest case, which is normally referred to as a demand-response "notch." The customer was operating at some expected level before the demand-response event, notches down to (or down by) a promised kW level for the agreed upon time, and then notches back to the expected level. If the buyer needed a four-hour block of power late one summer afternoon and the customer avoided the necessity of that transaction, one would naturally assume the customer should be paid appropriately. Then again, how close did the customer come to the promised behavior? After all, at the end of the day, this is about being paid for what actually happened and the buyer believing they received appropriate value for the demand reduction.

The information shown in Table 2-1 illustrates one such event. Notice that this customer's baseline is perfectly flat at 10 MW and that its pledged demand reduction was a constant 2 MW for four hours. The measured MW in the hours before the four-hour event were exactly coincident to baseline and the measured MW reduction during the four hours was precisely 2 MW in each hour. The customer then returns to the baseline immediately following the pledged period of time. Does this type of behavior occur in real life? Not often.

This behavior can be seen for certain types of industrial processes, but few others. In most situations, the customer's baseline is not perfectly flat, and there is some level of variation in that baseline from day to day. Said very simply, the confidence one can have in an individual customer's load changes against a baseline can be viewed in reference to some error band. If these errors are simply natural variations in day-to-day load shapes, one level of confidence can be held. If, on the other hand, these day-to-day variations are the result of operational decisions, other considerations might apply. So, let's first define what we mean by precision before considering some real-life examples.

Table 2-1
Notch-like Demand Reduction

Hour Ending	Baseline MW	Actual MW	Measured MW Drop	Pledged MW Drop
7:00 AM	10	10	0	0
8:00 AM	10	10	0	0
9:00 AM	10	10	0	0
10:00 AM	10	10	0	0
11:00 AM	10	10	0	0
12:00 PM	10	10	0	0
1:00 PM	10	10	0	0
2:00 PM	10	10	0	0
3:00 PM	10	8	2	2
4:00 PM	10	8	2	2
5:00 PM	10	8	2	2
6:00 PM	10	8	2	2
7:00 PM	10	10	0	0
8:00 PM	10	10	0	0
9:00 PM	10	10	0	0
10:00 PM	10	10	0	0
11:00 PM	10	10	0	0

What is Noise and What Isn’t Noise?

What does it mean to say that a baseline is repeatable? Is it simply good enough to use some mathematical average of the past “normal days” and possibly correct for weather effects? Today’s demand-response programs are struggling to varying degrees with this determination. Part of the reason they are struggling is they have not defined the underlying valuation of demand response. One of those valuations (or possibly one of the areas that diminishes value) is “noise.”

If you start at a theoretical level, noise is defined as random variation. Noise exhibits no pattern. Otherwise, the variation has a systematic nature to it, and that is not referred to as noise. So, let’s take a look at several load shape patterns for customers and evaluate their “noisiness” or lack thereof.

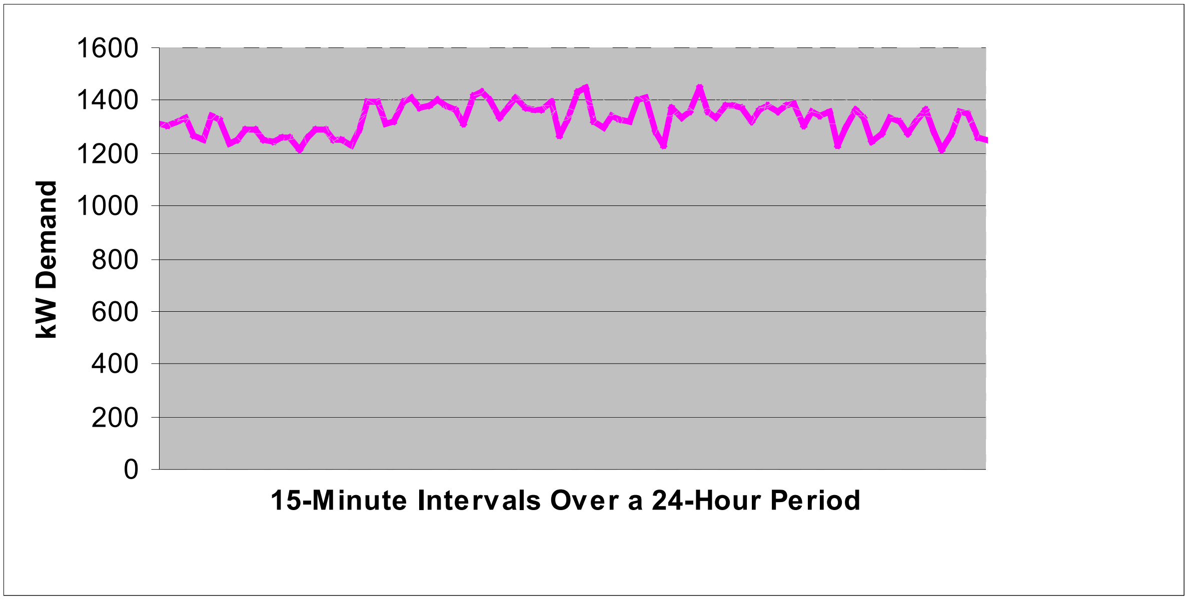

Figure 2-1
True Noise in Customer Load Shapes

Figure 2-1 is a sample customer load shape for a typical day. The data are 15-minute interval kW levels for the 96 intervals in a day. The data show a clear random (noise) component. While there is a slight systematic fall in the early morning hours, most of the variation is noise. Its random nature differentiates it from anything systematic in the pattern.

One of the most comforting elements about noise is that it is reduced as the time intervals over which it is averaged increases. For example, if we were to plot these same data using hourly averages (as the trading itself uses), the variations would be smaller. If we then averaged the hourly averages into four-hour trading blocks (the way trading is conducted in the Southeast on a day-ahead basis), we would further reduce this noise. And, if we aggregated customers of this type together, that too would reduce this noise component. So, noise itself isn't really much of a problem when settling demand trades unless the trades are for very short, near real-time use.

Many statistical tests can be used to analyze these patterns. However, tests that have been found to be most meaningful are a simple linear regression on the trading time period itself for the 15-minute data, along with a simple linear regression for the change in load from one 15-minute interval to the next 15-minute interval. This involves analyzing the load shape for the traded period along with the first derivative of the load shape for the traded period. This type of analysis of the first derivative of the 15-minute data indicates systematic changes (such as weather dependencies).

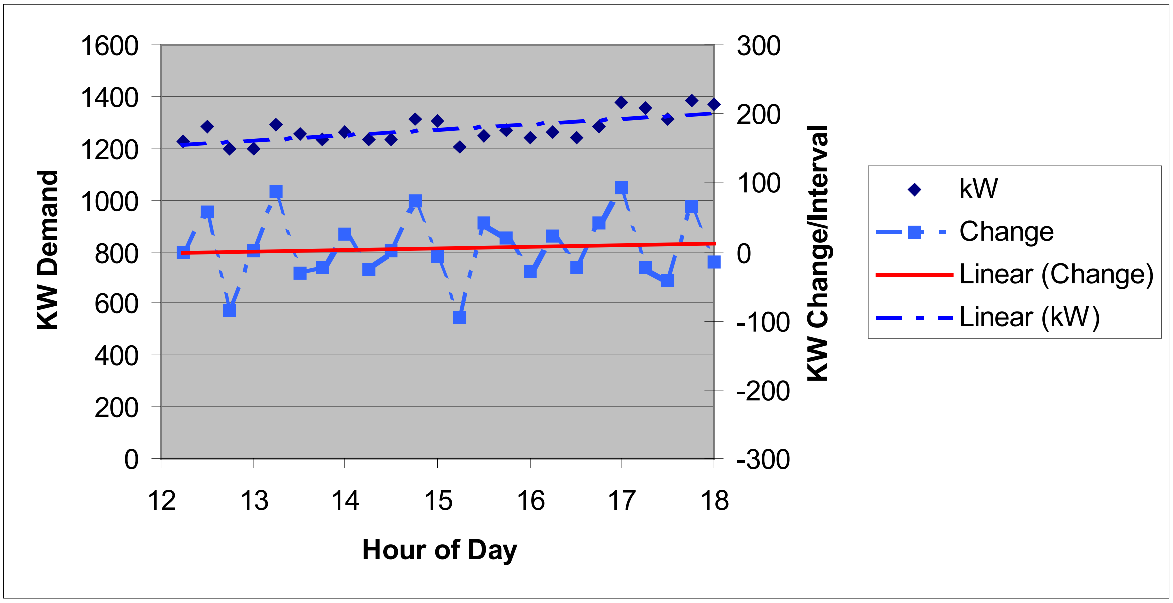

Figure 2-2
Evaluating Noise in Customer Load Shapes

For example, let's assume the trading period was from 12 noon to 6 pm on a particular day and we want to analytically evaluate the strength of the noise from the signal strength in the baseline itself. We do this by taking the 15-minute data, and then creating a second data string of the difference between successive 15-minute interval data elements (the first derivative of the data). Then, we use linear regression on both and compare the correlation coefficients along with the standard errors of the estimates. This is depicted in Figure 2-2. Now we have moved completely away from any subjective determinants and can compare customers to each other, as well as look at the consequences of aggregating customers together.

Please note that while most trading instruments focus on hours or blocks of hours, the evaluation of noise must be investigated at the 15-minute level to adequately determine systematic vs. random elements. The reason we emphasize this is the tendency for demand-trading data to be averaged into hourly sets. This is useful for settlement, but is less than satisfactory for informative evaluation of load shapes.

Completing this illustration, the noise in the example can be estimated by taking the standard deviation of either the load shape during the trading period (which turns out to be 54 kW in this case), or (if the small degree of rise during the period is to be ignored) by taking the standard deviation of the change in readings from interval to interval (which turns out to be 50 kW in this case).

The natural temptation at this point is to set specific trading rules as to the size of the pledged demand reduction in comparison to the baseline and these standard deviations and slopes. That can certainly be stated, but the reader will shortly see that simple rules are often simply wrong.

We have yet to establish the conformance metrics of the customer's amended load shape (as the customer takes its demand-response measures into effect).

Once you are satisfied with the intuitive and mathematical results from a review of the 15-minute data, and have assured yourself of the noise vs. systematic variations, it is time to examine this same customer's baseline and performance on an hourly basis. Averaging the 15-minute interval data over each hour cancels out much of the true noise and shows the systematic behavior more clearly. However, once again, do not assume we are recommending the use of hourly data to verify customer demand response. We are simply emphasizing how and why hourly average data tends to cancel out noise. Attempting to evaluate the noise in baselines will be frustrated by the use of hourly data. It is essential to acquire 15-minute data for such an evaluation. In fact, if 5-minute data are available from the meter, it provides an even better level of noise assessment. However, that is seldom possible unless the customer has an advanced meter-reading system.

Linear regression of the hourly, daily, or even monthly data against temperature (or heating and cooling degree days), and day of the week will then show how the systematic behaviors depend on weather, business cycles, and operating decisions.

When is an Average Meaningless?

We sometimes joke about averages and how meaningless they can be – for example, about the size of the average American family. But, averaging customer baselines and finding out they are meaningless for real-time transactions presents challenges. When is use of an average fair, and when isn't it? How do you define this in a way that both customers and market counterparties find defensible and quantifiable? As you will see, just because an average baseline seems repeatable from month to month doesn't mean it is representative of the load itself.

The obvious rule: an average is meaningless when it itself doesn't exist. The average roll of one die is 3.5 (the average of 1, 2, 3, 4, 5, and 6) but the number 3.5 does not exist on the die. Therefore, comparing a customer to a baseline that is an average of distinctly different operating conditions is intrinsically wrong. In fact, doing so will often result in a customer "gaming" the situation by bidding in what it knows to be a scheduled deviation from baseline. If you fail to detect that as being likely by examining the baseline in advance, you are bound to find out afterward and have customer service problems as you begin to enforce gaming rules.

The most common gaming situations that can be troublesome occur in the pulp and paper industry (where certain paper machines only run intermittently), specialty steel mills (that may operate one or more large arc furnaces and rolling mills), plastic extruder plants, etc. The only way to assess these customers is to take a close look at the 15-minute interval data and watch for "campaigns" (defined as several hours or even days of running a given line and then shutting it down to set up the next run) and other patterns of operation. There are statistical tests that can be run as well by looking at the frequency distribution of loads in a month or longer period. Most of the customer-operation problems will show as multimodal distributions rather than one normal-shaped distribution.

Standard statistical tests of significance can be used to test the confidence one can have in the underlying assumptions of multimodal behavior, but a simple "eyeball" scan of the interval data can save a great deal of time and effort, and avoid significant customer service problems later.

For example, let's assume a customer has a monthly peak demand of about 25 MW and a minimum monthly load of about 3 MW. Each 15-minute interval during the baseline-acceptable days in the month could be tallied into a certain number of "bins" of kW range. If we take each bin to be 500 kW wide, the first bin would capture all demand readings from 3.0 to 3.5 MW. The second bin would capture from 3.5 to 4.0 MW, etc. Plotting the number of observations in each kW bin vs. kW would yield a histogram of results. The results for this customer are shown in Figure 2-3. This is clearly "bimodal." In fact, given the broad dispersion in the mode to the right, it is entirely possible that this mode is a combination of two overlapping distributions. Please heed the warning to look at the kW vs. time plot depiction.

In any event, averaging the entirety of these data would be misleading and potentially lead to gaming on the part of the customer, especially if the baseline method selected is the highest X out of Y days.

Event-Day Conformance to Baseline

Traders are reluctant to trade something they can't measure and, left to their own devices, tend to force demand-trading partners into a liquidated-damages exposure for any failures to conform to pledged performance. On the other hand, most customers simply want to receive fair payment for what was pledged and have only limited control (in most cases) over the specific hour-by-hour loads. So, the balance here is clear. The customer wants to be able to offer a capability and the trader is looking to aggregate that capability for mutual economic benefit. Then, when the trader settles the net results against actual meter data, they want to be able to assign any consequential costs for lack of conformance to those who exposed them.

And, while generation agreements are intrinsically measurable, demand reduction is generally less so, except for the notch behavior (as discussed earlier) that only a few customers can offer. Compounding the challenge is the simple fact that this transaction only occurs when the economic opportunity in the value chain can be ascertained with sufficient lead time to allow both buyer and seller to commit to the transaction. This is one area where customers are intrinsically different than generators.

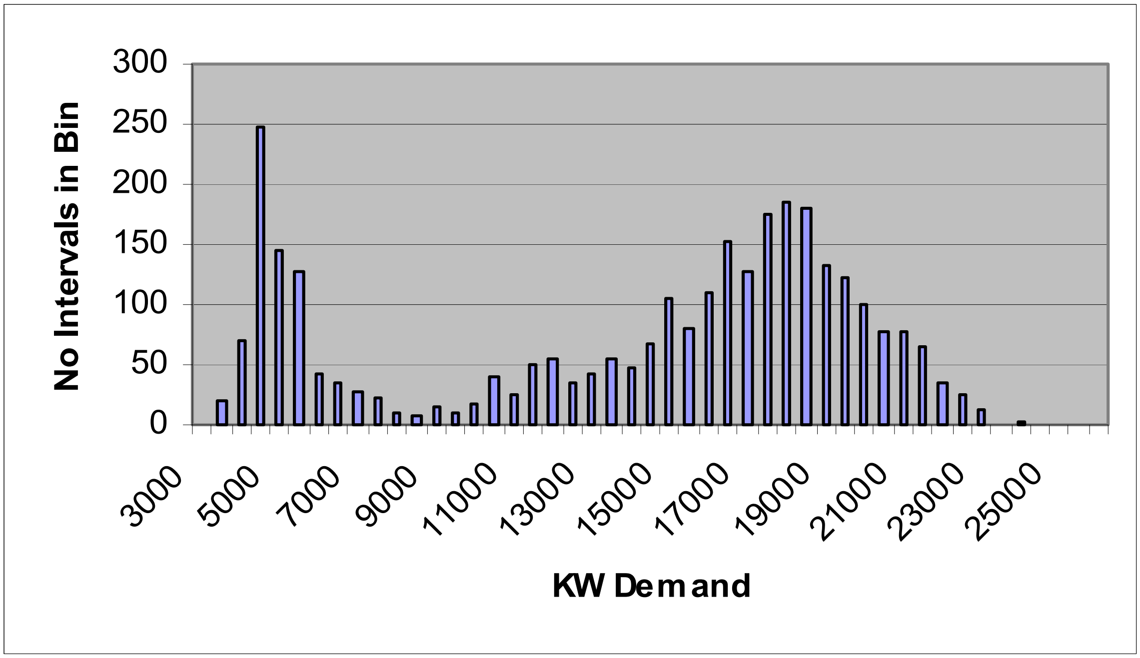

Figure 2-3
kW Demand Bin Analysis

Most customers prefer day-ahead (or even earlier) notification of a demand-trading opportunity. Obviously, the price for electricity in each hour of that event is known with less certainty day ahead than day of. In general, the price offered day ahead will be "discounted" by a trading organization to reflect that uncertainty. Customers who can wait until the last minute to commit their demand-response capability because it is automated into the market (such as with direct load controls or automated generator operation) can generally get a higher price for the same kW offered. However, many customers need a few hours or even a full day of advance notification. In fact, experience indicates that day-ahead notification strategies yield so much more resource (generally 3 to 5 times the MW and MWh potential of day-of or hour-ahead agreements) that they are financially worthwhile even though day-ahead price offers are generally lower than day-of and known with less certainty due to the nature of hourly price confidence.

Noise vs. Temperature Effects in the Baseline

One of the most compelling examples of how noise can be cancelled out is when complex loads are correlated against temperature (or the equivalent number of degree-days). If one attempts to correlate them against the temperature in each hour, the results may appear to have a significant correlation (even as high as 80-90%), but will prove to be extremely frustrating when attempting to establish a baseline in each hour due to the random noise component. However, when monthly total MWh and degree-days are correlated, the noise cancels and the temperature effects are clear. In fact, it is not unusual to get a 95-98% correlation using monthly data.

So, if we can get that close to a temperature relationship, why not use the monthly correlation applied to each hour using the actual temperature data? It is instructive to do that to compare the systematic and noise variations to the temperature-dependent variations. What you will see is that the baselines "shift up and down" and routinely depart from such a correlation. Actual knowledge of the customer in question would probably tell you why, but there are clearly factors other than weather influencing the customer.

Another warning is of importance here, especially for situations where the peak loads occur during the summer months due to cooling loads. There are two components to these cooling loads: the weather (and the amount of outside air brought into the building) and the internal cooling loads due to lights, operating equipment, and people. The former depends upon the weather and the latter depends upon the operating schedules. That is why we correlate monthly data with the number of operating day types to improve baseline predictions. Don't get too carried away with this, but the general rule for customers who do not exhibit markedly different Monday vs. Friday behavior is to keep track of the number of weekdays, Saturdays, and Sundays in the month (if they are different), along with the number of holidays. In most cases, two broad day categories will suffice: weekdays and non-weekdays. But, once again, there is no substitute for a comparison of day types before jumping to this conclusion.

So, with all this discussion about how demand response can be measured and variations attributed to the uncertainties of weather, operating decisions, and plain old noise, here are four general rules you can use to evaluate any demand-response mechanism.

General Rules for Evaluating Demand Response

Rule 1: Use 15-Minute Data and Not Hourly Data

While it is appropriate to use hourly averages to credit customer demand responses (since the wholesale market is usually settled using hourly averaged kW), it is better to present and evaluate individual-customer performance on an interval-by-interval basis (typically using 15- or 30-minute intervals). There are a multitude of reasons. The first is illustrated in Figure 2-4, where a customer could have either of the 15-minute interval data reads and still have the same average hourly load shape. While this diagram is hypothetical, it is actually rather typical of what you will see when commercial buildings are compared to manufacturing or industrial settings. Commercial building loads generally have lower interval-to-interval noise. As a result, smaller demand responses can be measured with confidence. In addition, by showing customers their interval-by-interval performance, the specific timing of any customer actions will show more clearly. Therefore, if a customer failed to curtail devices until 15 minutes after the hour, it will not get blended into an hourly average – it will show more clearly.

At some point, this additional detail can become unnecessary. Once customers have repeatedly performed demand-response actions, hourly reconciliation is probably adequate. However, even here it is always better to show the intervals themselves. And, if higher-level resolution is available (such as five-minute or better data), there is also the possibility of issuing alerts and alarms that certain conditions are not being met. Of course, this is also possible using 15-minute data as well, but is far better when five-minute or better data can drive the algorithms.

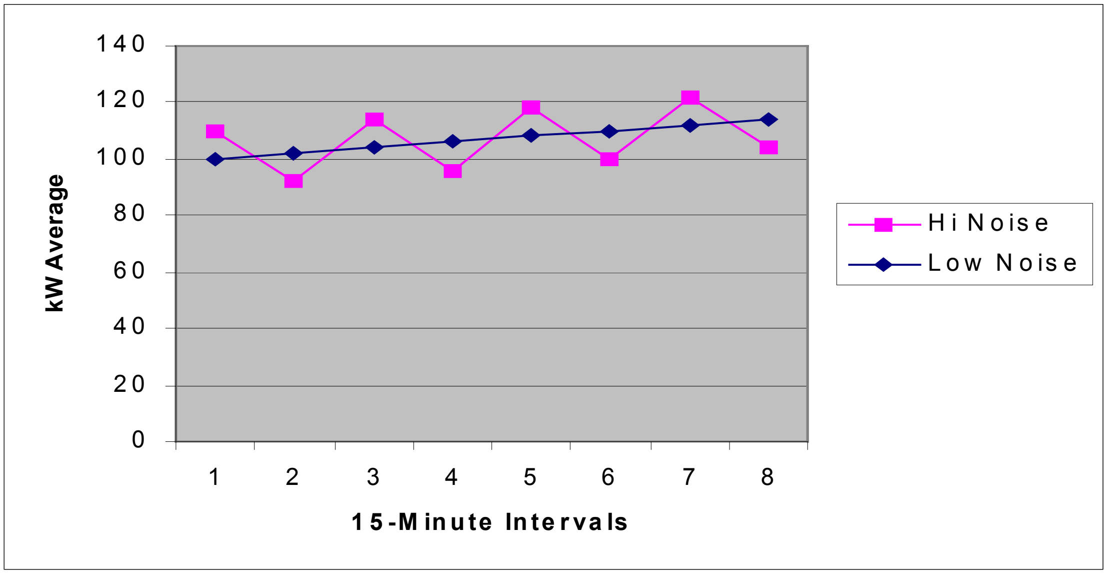

Figure 2-4
Comparison of Low- and High-Noise Load Profiles

Rule 2: Show Highs, Lows, and Averages in Each Interval

The second rule is to show customer demand response to all interested parties with the individual 15-minute highs, lows, and averages for each interval of the day in the baseline period (not the highest and lowest peak-day load shapes) as well as the event day superimposed on this same backdrop. Figure 2-5 presents an example of a four-hour curtailment in a commercial building using exactly that method as well as the effect of subtracting the event day from the baseline to produce a measured Net Load Shape Change (NLSC). This NLSC has proven to be the best measure to plot because it clearly shows any pre-event and post-event changes in load shape as well.

There are several very "comforting" characteristics of this specific customer's performance:

- The event day was slightly higher but quite close to the average being used for the baseline itself,
- The customer has a sharp ramp down to the curtailed condition as well as a rapid return to "normal" or average loads, and
- The period right after the event itself shows no apparent "rebound" or "recovery" effects that one might expect if the curtailment included interrupting a water heater or an air conditioning system.

Also notice that the NLSC shape is well defined and "notch like" and, in this case, very close to the pledged kW level of 125 kW for this customer. Building on this discussion, we can now define these metrics mathematically and see how a customer like this stacks up against a more typical manufacturing customer with higher noise and systematic as well as weather variations due to process refrigeration.

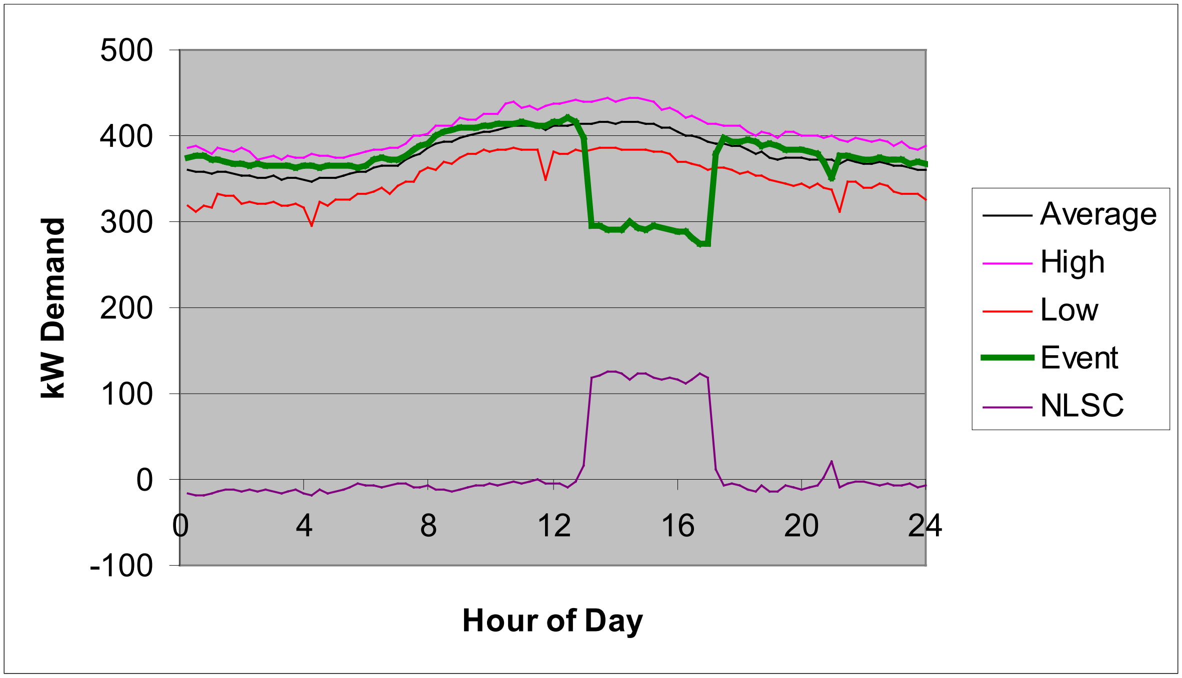

Figure 2-5
Suggested Event-Day Graphical Comparison Method

Rule 3: Quantify the Strength of the Net Load Shape Change to the Pledge

The demand-response notch can be characterized by area (kWh total for the pledged period) as well as shape conformance to the pledge. The latter can be easily defined using the typical performance criteria for liquidated damages of over and under performance against pledge. One might explain this to a customer as kWh promised vs. net kWh credited.

The typical commercial terms enforced where liquidated damages are used indicate that a minimum of 85% of pledge is expected in each hour and any "underperformance" below this level is debited against the credits at twice the credit price. This effectively reduces the net benefit and typically encourages customers to pledge conservatively. However, an energy trader can only trade what they know or believe they have, so overproduction is often debited as well. Common commercial terms indicate that up to 125% of the pledged amount is credited, after which no additional credit is offered. Said another way, the customer benefits are debited at the same price they are credited once 125% of pledge is achieved.

In the example shown in Figure 2-5, the conformance to pledge is extremely good with a pledged 500 kWh in the four-hour period (125 kWh per hour for four hours) and a measured 479 kWh, or a 95.8% volume match. In addition, the intra-hour variation against pledge never varied more than 10% in any 15-minute interval nor more than 6.4% in any one hourly period – well within any liquidated damages criteria. Plus, the periods immediately before and immediately after the pledge were close to baseline with extremely rapid ramping characteristics. This, by the way, is the typical load shape behavior of lighting system pledges, some pumping situations, and common when non-weather dependent loads are shifted to a standby generator.

While these aggregate statistics are an excellent way to summarize a customer's performance, they tend to be ineffective in educating customers about the specifics of how they performed and the impact of any liquidated damages on their net benefits. To do that, the type of presentation format of Table 2-2 (shown here implemented in an on-line demand-trading platform) helps customers and market counterparties understand the finances of pledge accuracy.

Table 2-2
Suggested Method of Presenting Event-Day Settlement

Estimated Settlement

Prices
Activity
Roster
Baselines
Settlement
Support
Logoff

Customer: Custom Manufacturing, Inc. - ATL#1 Curtailment Event Date: 2/11/2002
Account Number: 987654a

	Under Delivery	Over Delivery
At % Of Pledge	85%	120%
Debited % Price	200%	100%

Time Period	Offered Price Per kWh	Baseline kWh	Actual kWh	Measured kWh Reduction	Pledged kWh Reduction	Under Delivered kWh	Over Delivered kWh	Over/Under Delivered % of Pledge	$ Credited for kWh Reduction
12-1pm	$0.25	1,908	1,258	650	800	150	0	81%	$162.50
1-2pm	$0.25	1,879	1,059	820	800	0	20	103%	$205.00
2-3pm	$0.30	1,701	721	980	800	0	180	123%	$294.00
3-4pm	$0.35	1,787	767	1,020	800	0	220	128%	$357.00
4-5pm	$0.35	1,951	1,001	950	800	0	150	119%	$332.50
5-6pm	$0.35	1,894	994	900	800	0	100	113%	$315.00
			Totals:	5,320	4,800	150	670	111%	$1,666.00

Time Period	Under Delivered kWh	Under Delivery Correction?	Under Del. Debited kWh	$ Debited Under kWh Correction	Over Delivered kWh	Over Delivery Correction?	Over Del. Debited kWh	$ Debited Over kWh Correction
12-1pm	150	Yes	30	$15.00	0	No	0	$0.00
1-2pm	0	No	0	$0.00	20	No	0	$0.00
2-3pm	0	No	0	$0.00	180	Yes	20	$6.00
3-4pm	0	No	0	$0.00	220	Yes	60	$21.00
4-5pm	0	No	0	$0.00	150	No	0	$0.00
5-6pm	0	No	0	$0.00	100	No	0	$0.00
		Totals:	30	$15.00		Totals:	80	$27.00

Demand Reduction Savings: $1,666.00
Debits for Under Delivery: $15.00
Debits for Over Delivery: $27.00

Net Demand Reduction
Credited to Bill: $1,624.00

The label of Estimated Settlement in Table 2-2 is a nice way to alert the customer that this is the best guess about what the benefits are. There is always a possibility that there has been a meter reading error, or that there was some level of uncertainty in the final price offer (where companies guarantee customers that they get the minimum of the price offered or the settled hourly prices). In addition, there have been examples of ISO/RTO changes in those hourly prices after the event itself.

As indicated earlier, the best way to inform customers about how they are doing during a pledged event is to make their actual 15-minute interval data available to them during that time. Most companies that operate MV90 systems and read meters once a day can set that system to read the meters for those customers who pledged demand reduction once every 15 minutes and update this information on line. If they do not have adequate modem banks to perform this task, several third-party agents can provide this as a service.

Plotting those data against the historical highs, lows, and baseline average should show whether the customer's actual loads are in line with expectations. If there was any internal miscommunication or someone who was uninformed reversed the actions of the one who had performed the load reduction (e.g., someone notices that something is off that is usually on and turns it back on, not knowing that this was in response to a demand trade), this can be caught before it exposes the customer to a debit for non-performance.

It is also suggested that the settlement description shows the rules for implementation so that customers and market counterparties can understand the impacts of the rules. For example, Table 2-2 explicitly states that debits of twice the benefit are imposed for each kW below 85% of pledge and debits equal to the benefit are imposed for any amounts over 120% of pledge. The latter fixes the maximum amount this trading counterparty will pay for overproduction of pledged performance.

Obviously, the customer in this example did a good job at pledging and the debits were small in relation to the credits. If this customer had pledged the demand reduction and then failed to perform at all, they could owe money to the counterparty. This could potentially be quite a bit more than they could have saved if they had performed. This potentially embarrassing and painful situation can best be avoided through careful customer education and "testing" of actual customer capabilities before any demand trading actually begins.

In many cases, energy companies have signed up customers for demand-response programs without adequate regard for the consequential customer service issues that can occur. This happens quite naturally out of the bias to give "favorite" customers first crack at these opportunities without adequate regard for training, load shape evaluation, and actual demand-response testing. In addition, the tendency for customers to abandon curtailment and interruptible programs once they are actually called had many customers flocking to voluntary demand-response mechanisms without adequate consideration for the ways they are different.

Now, let's consider a typical industrial firm where the intrinsic noise is higher and there is some level of either weather dependence and/or systematic variation in the day-to-day load shapes. A common situation is illustrated in Figure 2-6. Notice that the event day is tracking up along the highest interval data records in the baseline period. This typically occurs when there is weather dependence in the load. While there is a clearly discernable "notch" effect as a result of the customer's curtailment actions, the use of a simple average for a baseline as shown seems intuitively wrong and intrinsically unfair.

The best way to characterize a comparison of the event day compared to the baseline is by plotting the NLSC as shown and "flagging" areas of concern. Flags can be set automatically in the automated settlement of these events, and customer settlements that do not set flags can be ignored while those accounts setting multiple flags might deserve further investigation and possibly even a customer visit to discuss matters.

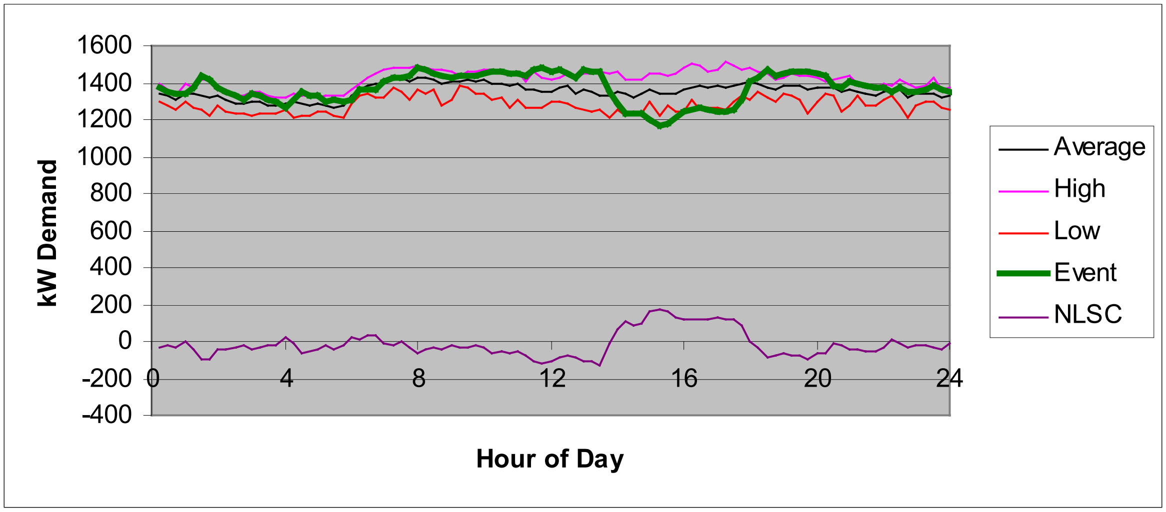

Figure 2-6
Graphical Comparison for Typical Industrial Customer

For example, one "caution flag" could be set when the customer seems to be deviating from baseline in the hours just ahead of the curtailment, and a second flag set for the period right after the curtailment event itself. A third flag could be set if the average load shape does not track the curtailment-day load shape well for the remaining hours outside of the pledge period. By the way, this third flag has proven extremely useful to correct baselines for temperature dependence. When the event day is consistently different from the baseline for the majority of the non-event hours, we find the easiest and fairest way to correct this is to "shift" the baseline up or down to reflect this departure. This is a substantially easier than attempting to do anything with temperature correction and has proven to be just as accurate.

Finally, you can compare the standard deviation in the NLSC during the non-event hours against the baseline to set a fourth and final flag. At this time in the transition of electric markets this fourth flag may be of little value. However, at some point in the future, customers will likely be charged for their use of ancillary services. When customers curtail noisy loads (such as arc furnaces), they may be entitled to some additional benefits beyond their hourly averaged demand reductions.

As you evaluate these noise and demand-response characteristics, you should rightfully be concerned about trading any customer demand response that is comparable in magnitude to the standard deviation in its baseline. Notice that we are not suggesting you draw the line for customer demand response as a percentage of its average or peak loads, we are saying that you want to compare the amount pledged with the noise in the baseline itself.

Experience operating demand-bidding programs indicates most energy companies insist that pledged demand response has to be a percentage of peak load rather than something based upon standard deviations of the baseline and day-to-day load shapes. For example, we hear some energy companies say that a customer's demand response must be at least 5% of its peak load. While there should clearly be some minimum demand-response level (such as the common 100 kW minimum for some number of consecutive hours) just to be sure there is an adequate response to justify the costs of measurement, verification, and settlement activity, it is often unfair to use a simple percentage of peak load as a rule of thumb.

Our statistical work indicates that demand trades should be greater than at least two standard deviations of the noise in the baseline. This can mean a minimum demand-response requirement may be 20% of the peak load for some intrinsically noisy customers. Of more interest and use should be the flip side of this argument. Our experience indicates that the standard deviation for most commercial customers is so low that demand trades of as little as 2% to 3% of the peak load can be traded with almost absolute confidence.

If the customer shown in Figure 2-6 had bid the same 125 kWh for each of four hours as the customer illustrated in Figure 2-5 it would have achieved that demand reduction. However, it is likely that this customer might well have argued that they deserved more. The easiest way to anticipate this viewpoint is to notice and inform the customer that the NLSC curve has significant "negative effects before and after the pledge period."

Before concluding that this is some form of weather dependence, take a careful look at the situations shown earlier in this chapter. If the daily and monthly kWh consumption at this facility correlate well with weather indices, there would be a basis for correcting the baseline with temperature or the corresponding degree-day determinations. However, remember to perform these correlations on a monthly basis to average out the random noise. Moreover, always consider whether the customer service and economic valuation of any customer's demand response is worth the trouble to deal with this level of detail. It is always better to avoid trading with a less-than-desirable counterparty than to attempt to clean up the mess once you do have trading disagreements.

There is a natural tendency on the part of the trading organization to feel somewhat self-assured about the impacts of liquidated damages on this type of account. One can almost hear them saying, "See, I told you these customers are fluff!" Clearly, this customer's baseline is noisier than the prior one, and the liquidated damages criteria probably are adequate to cover the financial uncertainty associated with this customer's performance. However, there is almost certainly going to be some level of disappointment when viewed from the customer's perspective. That is why we must highlight the fourth and last rule.

Rule 4: Always Run a "Test" of Customer-Claimed Capabilities

It is not clear why the obvious step of running a test is so often overlooked. Most customers do not know what they can pledge until they have attempted to perform such a test. This test not only helps the customer communicate required actions internally, it also illustrates the actual kW and kWh performance they should expect. This is also critical when the energy company is asking the customer to perform against a block pledge as opposed to an hourly pledge.

For example, Figure 2-7 presents an hourly load shape analysis for a commercial customer with a standby generator who pledges to completely "disappear" during a four-hour curtailment event. Notice that it shifted its loads to the generator slightly before the beginning of the first hour of pledge and then shifted the loads back to the serving energy provider slightly after the hour ending their pledge period.

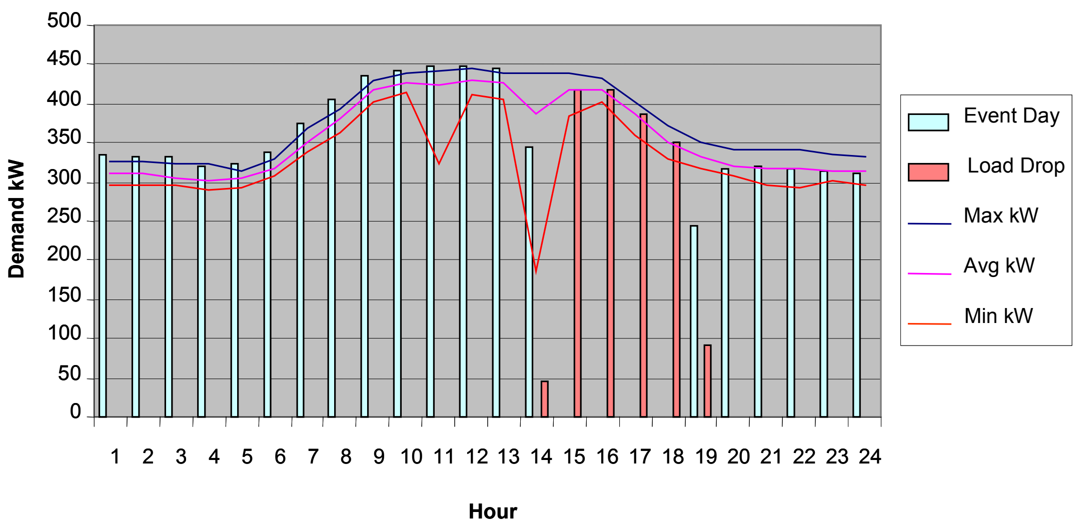

Figure 2-7
Commercial Customer with Standby Generator

First, notice how "tight" the range is for the maximum, average, and minimum kW demand in almost every hour for the baseline period. Next, notice the event day is reasonably consistent with historical data and generally falls within that high-low band. However, also notice that this customer basically peaks around noon (even before the pledge period). Finally, notice that while the day-to-day variation in load is generally small (a relatively narrow max to min band for most hours), the load is declining quickly during the pledge period.

This is not an unusual customer situation, and one that requires extreme care during the customer acquisition and education phases. Unless care is taken, this type of customer is almost certain to miss-pledge and become a customer service challenge when its actual performance against pledge is verified and settled. They are most likely to miss-pledge due to one or both of the following errors:

- They pledge the nameplate size of the generator failing to have even looked at the loads displaced when the generator actually operates.
- They pledge one amount for a block of hours and thereby fail to conform to that pledge due to the intrinsic characteristics of the load shape itself.

It is not at all unusual for commercial customer loads to fall off during the afternoon. This can make it difficult for this type of customer to pledge a block of hours at the same kW level, which is the common demand-response pledge requirement in the Southeast at this time. After all, what should this customer pledge for a four-hour block? The average kW (something close to 390-400 kW) is probably reasonable. However, why should the customer potentially be penalized for over and under delivery given they are performing to the fullest extent possible?

This points out one of the basic tenants of how demand-trade valuation is communicated and we will revisit this topic in a later chapter for this specific reason. It is entirely possible, and in fact likely, that one counterparty would want exactly what this customer can do to reduce demand while another might not and would, in fact, penalize the customer for its performance.

A Generalized Pledge-vs-Actual Performance Model

The final step is to generalize this characterization of load shape change and codify it into metrics for commerce. A baseline is determined to be fair if it reflects what the customer, on average, would have had as a load shape without economic incentive. The offer to pay for demand reduction produces a net load shape change.

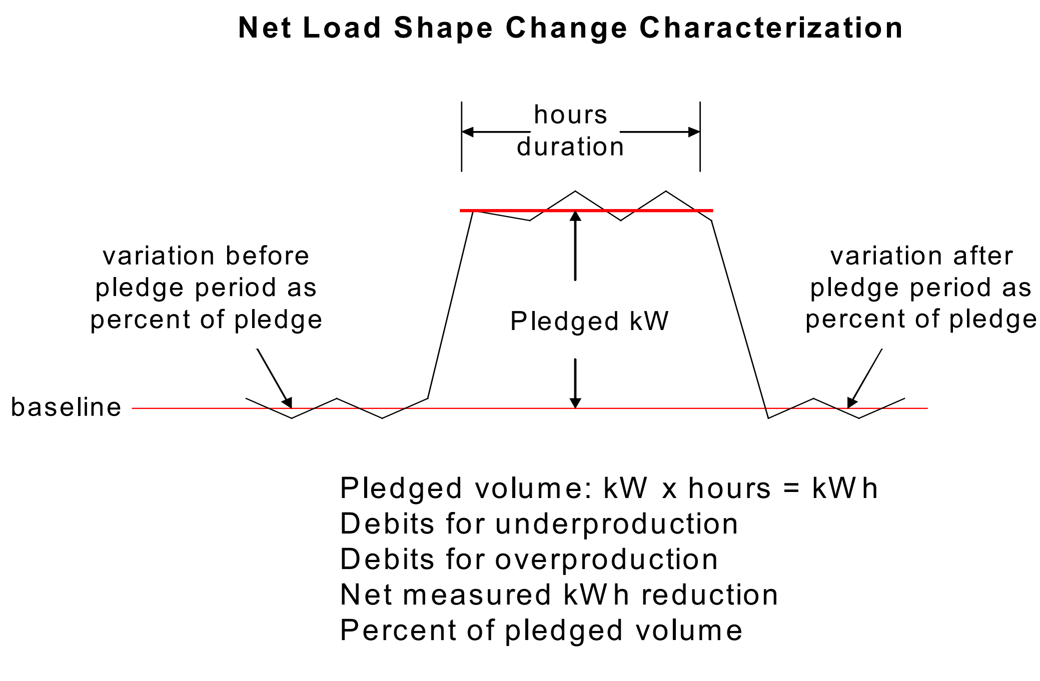

Figure 2-8
Suggested Standardized Net Load Shape Characterization

Payments are based on the assumption that the customer will consume at the baseline level up to an appropriate period of time before the pledge period and then meet the pledge within some acceptable variation. If the customer deviates from the pledge by more than the prescribed level, debits are taken against any credits given. The illustration shown in Figure 2-8 is for a constant pledged kW for the entire pledge period. Obviously, if prices vary in each hour of the pledge period it is likely that the customer demand response will vary as well in relationship to those price offers. In each case, the customer's reduction against baseline is credited following the general approach shown earlier in Table 2-2. The only missing element is some level of description about how "noisy" any one customer's baseline is in relationship to others. There are several reasonable statistical descriptions that are fair and informative, such as standard deviation in the baseline itself, as was presented earlier.

Obviously, the result being purchased is an improvement in net load shape. That improvement may be purchased as a block (e.g., 4 hours at the same kW level), or by individual hour, but the result being purchased is a reduction in hourly loads against some predetermined value. Some buyers and sellers will be willing to agree to four-hour, single-value deals while others may insist on hour-by-hour reconciliation. Commercial terms along these attributes are possible, can be generalized, and will be considered in Chapter 3 of this document.

3
VALUING DEMAND RESPONSE AND TRADING ACROSS TRADITIONAL BOUNDARIES

As explained in Chapter 2, demand trades in the day-ahead and hour-ahead markets can be numerically characterized by their average hourly values and metrics of conformance to a requested net load shape change. For example, one buyer of the resource may wish to see the same demand response level in each hour offered (typically a four- or eight-hour block) and have very tight bounds about how much the customer can deliver over or under its commitment. Another buyer in the same region may have much less sensitivity to the hourly variations and be willing to accept any load reduction offers at its price just as long as there are at least so many contiguous hours of demand response (typically four or more contiguous hours).

As one might expect, it is entirely possible that regional energy companies might differ dramatically in their valuation of demand-response resources. For example, hydro-dominated systems, or those with adequate capacity and significant hydro capability may value demand response only as energy. Systems dominated by thermal power plants may value demand response primarily as capacity and an offset to the equivalent of demand charges in commercial supply agreements. Others may only want demand trades for emergency situations, especially if they are “long on supply” or feel regional energy markets are either too unpredictable or complicated to involve their customers with.

Demand trading offers a useful way for customers to act as market counterparties to assure all regional energy providers that prices will be held in check by customer demand response. However, the multiple relationships along the electricity value chain complicate the ability of potential buyers (other than the customer’s normal energy company) to gain access to the demand-response resource. How can multiple offers be made to customers from beyond the normal energy company’s boundaries and be compared with due consideration of all the rightful costs and risks of the market participants? For example, if there is an independent curtailment service provider and the serving wires company was otherwise meeting the customer’s energy needs, the wires company could find itself seriously overestimating the load to serve and paying significant balancing costs. In addition, the incumbent may have a fuel clause or some other disincentive for trading the customer’s demand-response capabilities. In many cases, the wires company would probably prefer to manage its own affairs without any disruptive elements from others in the region.

A pre-requisite condition, of course, for this type of trading beyond traditional boundaries is that the customer has to be permitted to see offers from (or make offers to) entities other than its traditional energy company. This permission requirement would seem silly in a fully deregulated market, but is a necessary step in the partially deregulated one we see across most of the U.S. at this time.

For example, if the customer has an existing agreement with its traditional energy company for curtailment or interruptibility, the energy company must first be sure it does not need that capability for native requirements. After all, it has a rightful first call on that customer option.

However, what if the customer can exceed the existing agreement's requirement by voluntary action (which most can)? Most customers can provide significantly more demand response than the curtailment agreement, but only on occasion, and possibly not without significant advance notification. In fact, there are some customers who can completely re-arrange their production schedules a week ahead to change their load shapes for one or two days. How can such offers from customers and from energy market counterparties be matched? That is the main subject of this chapter.

Price Transparency Can Be Disruptive

The energy company can be faced with an intellectual and relational dilemma. By exposing customers to offers from others, it may face difficulties with the perception that any prices it has offered are "fair" and the possibility that it may experience "agreement migration" in which a customer jumps to a different program from another energy company. For example, let's say the energy company has a curtailment agreement with customers that discounts their rate by $2 per kW per month for all kW reduced, and that the call on that kW can be for up to 100 hours a year. On average, that amounts to a payment of $24 per kW per year for 100 hours of use, or $0.24 per kWh ($240 per MWh). That may seem fair, but consider the fact that the customer discount is not tied to operation of the program – it is being prepaid.

Therefore, the customer's perception is that it does not get paid for what it actually does, but instead it as been paid for the promise that it will perform in the future. Then, one day during which the incumbent energy company doesn't want to call on the customer, the customer is provided with an offer from another market counterparty for $0.50 per kWh for a given block of hours. This could seem like a better deal and the customer might ask to get off the involuntary agreement thinking they could see more annual value.

Now, let's set aside any contractual limits on this and stay hypothetical just to look at the issues in the broadest of terms. You might rightfully assert that there is no assurance of annual benefit when you simply wait around for offers from third parties like this. You would have plenty of facts to support your position. In fact, many Midwest energy companies were faced with this fact when the summers of 1998 and 1999 produced significant price spikes that interested most of their customers in voluntary agreements, only to be followed by the summers of 2000 and 2001 where there were no price spikes. Naturally, many customers then asked to be put back on the involuntary agreements. (On one level, this is no different than the supply-side tension between signing forward agreements and knowing you will at least be able to run your power plant versus waiting in the wings for seasonal spot market price opportunities.)

So, what is the right way to address this situation? One could design the agreements carefully to discourage gaming and unreasonable switching. In addition, this situation could also be used to educate customers about market operations and the reasons behind the different types of rewards for different types of programs.

In situations where the customer's demand response will be traded beyond the traditional boundaries, the serving energy company could implement a "permissive switch" approach. Such a switch is normally off and can be turned on to permit others to offer price signals to a customer. Permissive switches are common in control systems to be sure that prerequisite conditions are met before an action is taken. The reader might be interested to know that many demand bidding programs have this switch if for no other reason to avoid confusing customers about when they can respond to outside offers.

While some free-market advocates will argue that the customer's demand response is its own and it has a right to trade the demand response as it wishes, others will disagree with this position. It is true that letting an energy company have this freedom to set a permissive switch could limit the resource availability. Nonetheless, many will insist that it is simply a fact of life at the moment in the transition toward open, competitive electricity markets.

Presenting Multiple Offers to Customers

What would it look like to present multiple offers to customers? Figure 3-1 shows one manifestation of this capability for a summer trading season where the high prices generally occur in the late afternoon hours. Notice that the traditional energy company (LSE) and another energy company (Bid offer A) are each offering variable prices in the hours while Bid B is for four hours only at one price.

Table 3-1 presents the criteria for each of these bids that the customer must consider in responding to these offers. The first item most customers will consider is the total economic benefit for a kW of demand reduction for the hours offered that day. The example here indicates the LSE is offering the highest total value for the period noted but the offer from Bidder A is a near second in that criterion.

The next topic in the list is the potential imposition of liquidated damages (LD) and the specific nature of how they are implemented. What we see here is that the LSE and the offer from Bidder B have no LD. The design of the LD for Bidder A involves a penalty of 200% of the offered price for each kW that is under 75% of the pledged kW amount. It also has a penalty of 100% of the offered price for each kW that is over 125% of the pledged kW amount. These LD terms are more generous than what we generally see, and what is considered reasonable by most customers who have access to meter data more frequently than once a month.

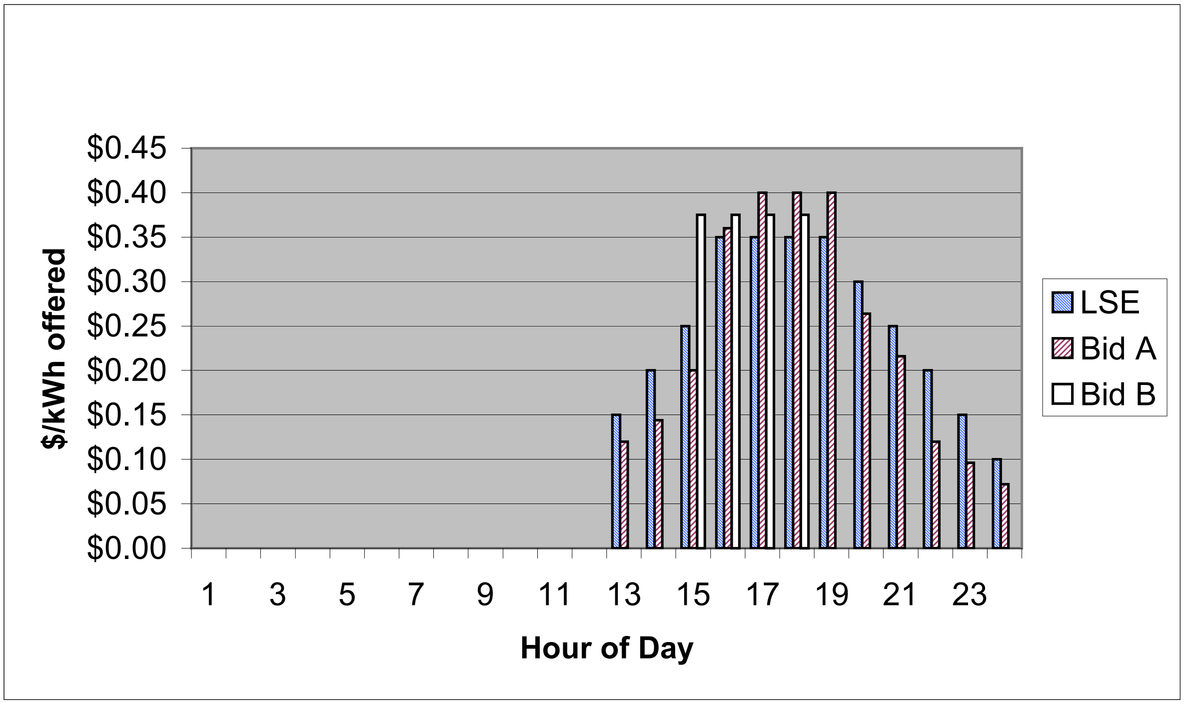

Figure 3-1
Demand Trade Bid Comparison for Typical Summer Day

Table 3-1
Example Demand-Bid Criteria (Summer)

	LSE	Bid A	Bid B
Economic Benefits ($/kW/day)	$3.00	$2.79	$1.50
Are LD Imposed?	No	Yes	No
% Under Pledge for LD	na	75%	na
% LD Debited for Under Pledge	na	200%	na
% Over Pledge for LD	na	125%	na
% LD Debited for Over Pledge	na	100%	na
Must Pledge All Hours?	No	No	Yes
Must Pledge Min Hours?	4	6	na

Finally, the notion of the minimum number of hours the customer must pledge is shown. All of the attributes in Table 3-1 can be automatically enabled on an exchange so that customers agree to terms and that bids conform to requirements.

The same situation can be presented for the winter season as shown in Figure 3-2.

Most energy companies have two peak periods to contend with in this season: the early morning and the early evening. These peaks can rise quite rapidly when electric resistance water and space heating are widespread in an area and the weather is normally mild. Energy companies in Florida are quite familiar with this and normally have direct load control switches on many of their customer's facilities to provide some level of control. The example shown here is more typical of energy companies that experience cold weather most of the winter. Notice again that the LSE and two other bidders are presented.

Table 3-2 presents an example of a common set of attributes for the winter season. Once again, any requirements on the part of the buyer to impose liquidated damages or block-bid constraints are presented and the customer bid response can be automatically constrained to conform with bidder requests.

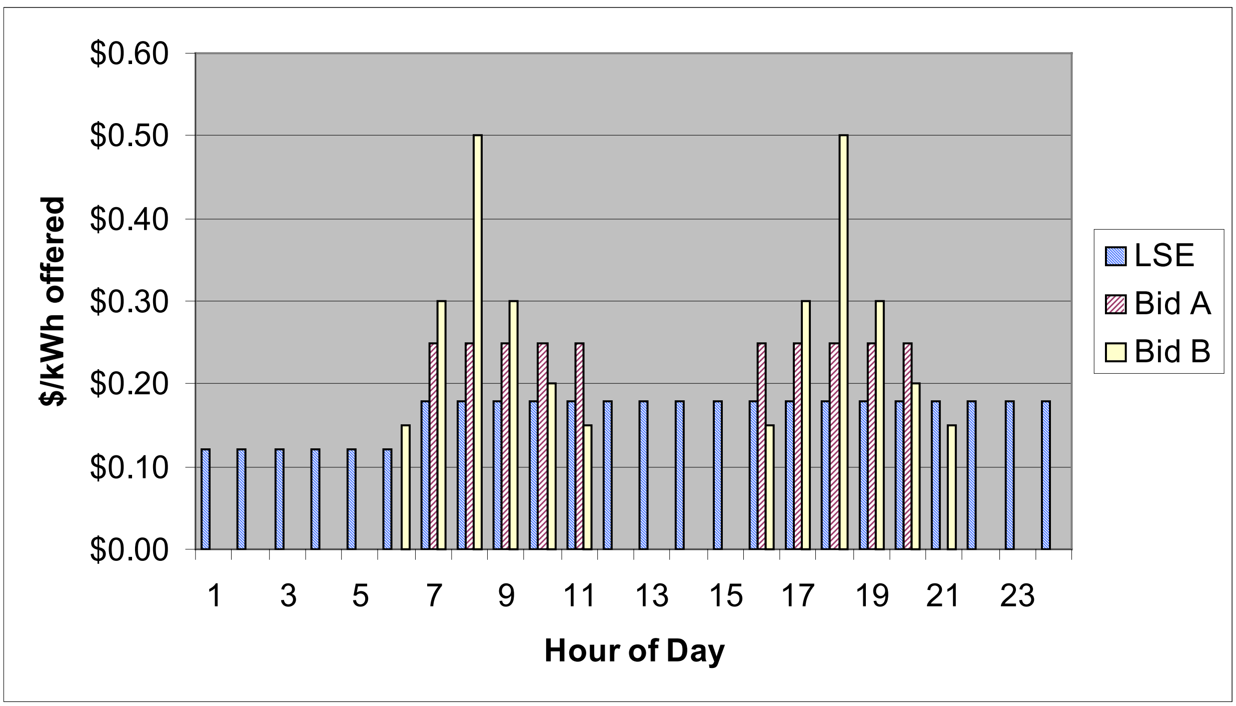

Figure 3-2
Demand-Trade Bid Comparison for Typical Winter Day

While this all sounds elegant and smooth flowing, there are lots of things that can go wrong in this model. First of all, just because the customer responds with an offer to reduce load doesn't mean it will be accepted. The price-offering bidder may find they cannot get enough demand response to meet their needs. Or, they may get an inordinately high response and have to accept only a portion of the demand reduction that is offered.

Table 3-2
Example Demand-Bid Criteria (Winter)

	LSE	Bid A	Bid B
Economic Benefits ($/kW/day)	$3.96	$2.50	$3.20
Are LD Imposed?	No	No	Yes
% Under Pledge for LD	na	na	80%
% LD Debited for Under Pledge	na	na	200%
% Over Pledge for LD	na	na	125%
% LD Debited for Over Pledge	na	na	200%
Must Pledge All Hours?	No	Yes	No
Must Pledge Min Hours?	8	na	2

So … Who Gets to Play?

As in virtually any market, there will often be an imbalance between the number of buyers and sellers, and a resulting price shift away from equilibrium. Markets should permit the ready matching of buyers and sellers, but the typical size of the demand trade is too small to be considered individually as an alternative to supply, and therefore some level of aggregation is usually needed.

In today's markets, the most common buyer is the ISO/RTO which will take the offer exactly as offered and is not trying to fill a given block size (as long as it is more than one MW), so this problem of deciding who gets to play goes away. However, as the regional energy markets start to use customer demand trades within the forward bilateral positions, this will be a potential issue and deserving of thought.

The regional bilateral electricity markets generally avoid "odd lots" and try to standardize on 25 MW and higher block sizes, and most often 4- and 8-hour blocks of hours at that level. There are many ways that blocks of that type can be filled:

- First-come, first-served basis: literally in the order in which demand-response offers are received.

- Batting order: those whose bids are accepted move to the end of the line to let others have their turn.
- Price-based bid stack: customers offer demand response at a price and a clearing price for the block size is calculated.
- Integrated into ISO bid process.

All of these methods have advantages and disadvantages. The first-come, first-served model tends to be difficult when there is great disparity between customer demand-trading capabilities. For example, imagine you need 25 MW and have five large customers each capable of 8 MW of demand response and 50 customers capable of 1 MW each. You place an offer and the website accepts bids. When you get up to the number 18 MW, the system will start rejecting your large-customer offers. You know what that will do to your ongoing relationships and participation.

The batting order strategy corrects part of this by making sure each customer gets a turn at bat by placing any rejected customer bids from one day at the head of the batting order for the next. You can certainly see how this can create customer service problems when the prices offered vary significantly from day to day.

It should be noted that customers will not respond to the same price the same way on any given day. Manufacturers become more inelastic to price offers as you get to the last days in the month if they are having difficulty meeting production quotas. On the other hand, some manufacturers become more elastic over time if they have met production and were considering producing to build inventory. Customers have been observed to reject prices one month at $.20 per kWh and take offers at $0.10 per kWh the next month due to this.

One could always expand on what California did with its demand-bid program where customers indicated a price and a load they would be willing to reduce in each of three four-hour blocks. The price points were pre-selected and a customer could offer progressively higher kW at the higher price points. Then, the buyer for the Cal ISO could simply select the clearing price and all loads committed at or below that price would be selected. Ironically the Cal ISO system involved an “as bid” approach, where the customer only received the price they offered (and not the highest cleared price in the stack).

There is no question that the Dutch auction process is the one that brings forward the least-expensive resources if the market is not being manipulated by withholding capacity, etc. While any market (e.g., California) can be manipulated through loopholes in the rules, we must handle markets with appropriate rules and market surveillance mechanisms. There’s no need to do away with the Dutch auction – we just need appropriate rules for the way resources can be bid into the auction.

Bid and Counter Offer

Flexibility can be an important factor contributing to improved program performance in demand-bidding programs. One example of a more-flexible approach is a simple bid and counter offer model that customers seem to like.

For example, consider the situation where the trader notices that the customers are considering the offers because the trader can see that they did go to the web site to check out the offer, calculated how much they might save, but then did not submit a pledge because the price was apparently not quite high enough. Traditionally, participants in voluntary price-responsive online demand-reduction programs either receive a price from their energy company or they submit a price of their own determination. In either case, the receiver of the price had a single option: take it or leave it. If the receiver of the price found the price too low or too high, the only choice was to not accept it, with the result being no demand response from that particular customer. While this model certainly remains adequate for the majority of demand-response programs currently enabled, in some cases more flexibility – a negotiation step – could be helpful.

The answer is to allow the receiver of the price to counter with a price of his or her own, whether the initial receiver is the customer or the energy company. This ability to counter-offer the original price solves a number of potential problems and adds a great deal of flexibility on both sides of the transaction, with the primary result being an increased likelihood that demand response is delivered.

Another technique is to take the identical approach from the opposite direction that works well for the largest customers. This would ask the customer to “bid-in” both the load reduction and the price to the trader. The trader can now counter-offer the price from the customer. The trader can hold the load pledged constant or ask for more or less at the countered price. Once the customer receives the counter offer he or she can accept it or reject it. In the case of a rejected counter offer, the customer can submit another bid. This is starting to look a lot like a real market at work!

Market rules are extremely important here to make sure the auction is fair to all. In general, the highest clearing price offer should be given to all who bid at or below that price to bring forward the most resource at the lowest possible price. If there are multiple price offers and any significant disparity, the customers are bound to find out and the fairness of the auction will be challenged. In fact, it is also deemed unfair by state PUCs in some parts of the country that any customer demand-response offer at a price stated by the utility is denied (just because the utility received more demand response than it needed). In those jurisdictions, the energy company must accept all bids or reject all bids.

How Accurately Can You Predict the Displaced Price in a Stack?

Does anyone really know what demand response is really worth for any hour or group of hours? It is important to take a close look at how difficult this question can be before implementing a program. Let's start out with very simple situations and look at how demand trading can be implemented to make the economics work.

Today it is common to find fully hedged energy service providers with source agreements covering all of their needs. These are typically referred to as full requirements agreements. Some portion of that will typically have demand and energy parameters for the last MW and MWh purchased. Let's make the math easy by assuming the agreement has a demand charge of $6/kW per month and a 100% ratchet on that for the next 11 months. So, the last increment of demand purchased on that agreement for one kW used only one day for four hours would cost $6/kWh times 12 months or $72.00/year divided by 4 hours or $18.00 per kWh ($18,000 per MWh). Avoiding the use of that demand for such a short period of time is clearly cost effective, so how can we do that?

We could simply pre-arrange a call option for customers' demand-trading capabilities at a number less than that. Let's say it was half of that or a reservation fee of $3/kW per month. The terms might be up to 5 calls in any given month and up to 15 calls a summer season with some level of hourly exercise price to cover out of pocket expenses such as the operation of a local generator. The concept is to displace the agreement in question by guessing the days that would have otherwise set a new peak. This concept is, by the way, in common use in many areas of the United States for demand trading and it is usually successful if you can acquire about 20 customer actions for up to four to six hours during an event day.

What we are doing is clipping the sharpest section of the curve representing kW vs. the number of hours at or lower than that kW in a year, also known as the "load duration curve." A typical example is shown in Figure 3-3.

Notice how steep the curve is as demand trading is first engaged. Each successive demand-reduction level has the likelihood of more hours of actual use. And, as the number of hours required goes up, the likely incremental value to the customer goes down. This law of diminishing returns is also problematic in that it is entirely possible to displace so much demand that the price-offering company begins decrementing another agreement. The problem can best be summarized by stating that the price offered to customers can only be made accurately if the volume impacts of their demand-response changes were factored into the price offer itself. It is this fact that causes many economists to insist that the cleared price come from an ISO/RTO.

This may sound a bit strange until you take it to an extreme. For instance, if an energy company had only customers who were completely demand responsive, they would never face high prices in the market and could offer regional counterparties their customer's demand-response capabilities. But, if that capability was so great that it killed the price spikes for the entire region, how could that energy company know the fair price to offer in the first place?

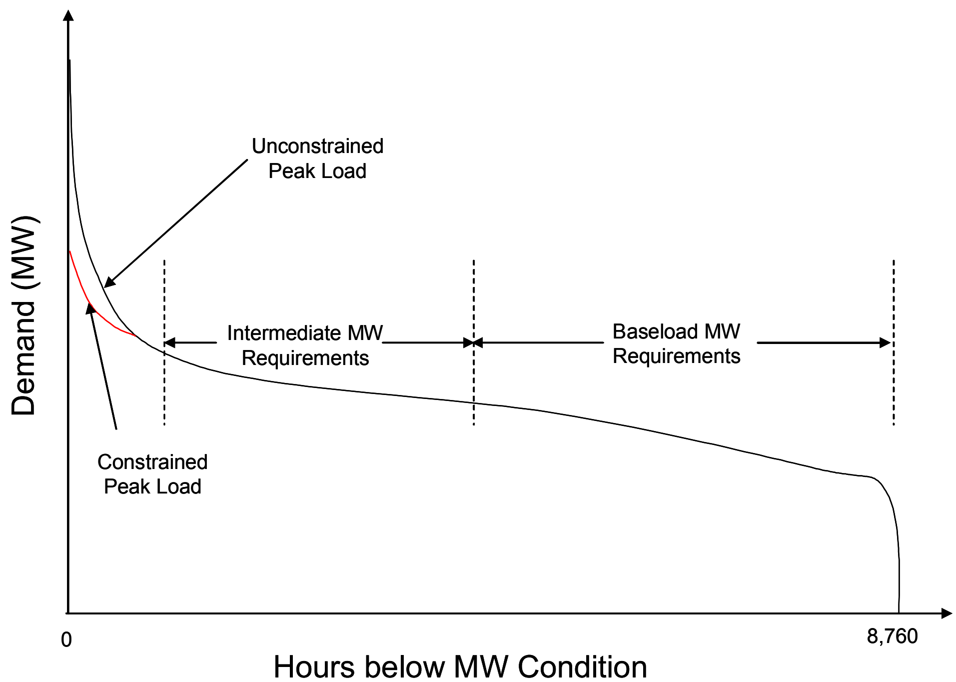

Figure 3-3
Call Option Impact on Load Duration Curve

A theorist would simply say this is no different than making a decision about how much insulation to put in the attic of your home. You would keep blowing in insulation until the added cost would no longer provide an adequate return on investment. But, you don't know how much you save in regional energy markets. This is one of the primary reasons to allow customers to see multiple offers – it is better to let the market provide price transparency, rather than having any one company attempt to reason this through to the satisfaction of others.

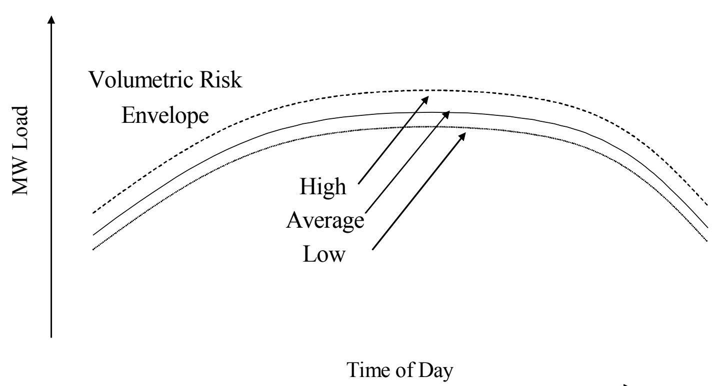

Figure 3-4
Week-Ahead Volumetric Risk Envelope

The complexity of the actual book of trades held by any one company to serve its customers makes this determination even more difficult since there seldom is just one agreement in play. Companies that trade in the spot markets have day-ahead and day-of agreements for each hour and may have a varying comfort level of where they are in the market for each hour of the day. For example, let's take a closer look at how this perspective can change. Figure 3-4 shows a hypothetical energy retailer who looks out a week ahead at the load they will have to serve in each hour of each day of the week. There is some probabilistic range to this forecast, but it can be bounded with a reasonable level of confidence using today's analytical tools.

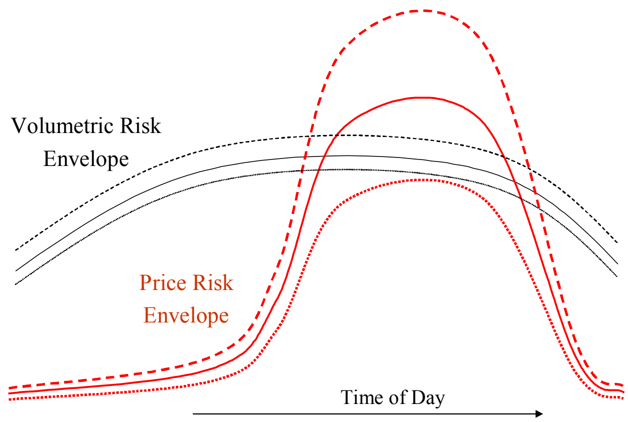

Figure 3-5
Day-Ahead Price & Volumetric Risk Envelope

As the time lapse between planning and real time diminishes, the expected range in load variation that will actually be served tightens up. At some point along the way, the energy company can begin the evaluation of its risk exposure in both price and volume. For example, consider Figure 3-5, which shows the volume and price risk estimates for the day in question. The most common action point for the first correction is day ahead, although we have seen two-day-ahead trades as well. Balance-of-the-week and balance-of-the-month demand trades have also been executed, but they are far from the norm. The price estimates are much more variable during a few hours of the day than the variations in volume expectations. But, what is more important to consider is the variation in forward position during the high-price risk period.

For example, let's assume the serving energy company was fully hedged with previously arranged supply agreements as indicated in Figure 3-6 for all of the high-priced hours. Depending upon the scale of this diagram, there might even be a block of power they might agree to sell off into the spot market on a day-ahead or day-of basis. The energy company takes significantly more risk day ahead than they take day of, but they will get more customer response day ahead than they will get day of. And, sure enough, there is very little volume or price risk as you get to within an hour or so of real time. Unfortunately, the number of customers who will participate is down as well.

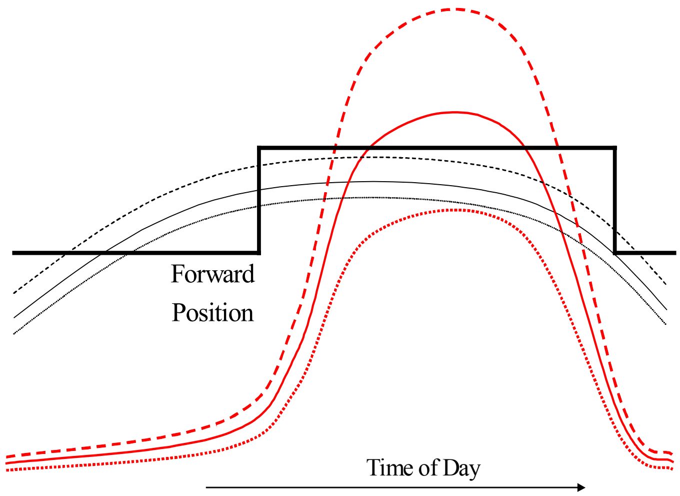

Figure 3-6
Forward Position Effects on Risks

Attitudes and perceptions of risk change depending upon the forward position the energy company has in relation to price and volume exposures. Where the spot prices are significantly above retail rates and the energy company is not adequately covered with prior forward positions, the combination of price and volume can be devastating. If that company has any advanced notice that this is likely and has a relationship with its customers for demand trading, the obvious will happen: they will work creatively to get customers to use less energy during this time period and in some way share the net benefits.

However, if the customer's demand-response actions cause that energy company to displace not only the spot market transactions but also dig into other prior forward purchases, the net economic benefit can be destroyed by overpaying for some of the demand response. Therefore, the net position before and after the demand trading must be assessed if the energy company is going to feel right about any offers.

The problem is compounded even further by the fact that the book of supply-side trades is not easy to visualize and traders are very busy people. There are buy and sell agreements in the book and it is not at all clear what the net displaced transaction will be in any given hour on a day-ahead or even a day-of basis. Remember, these books are not static spreadsheets. There are buy and sell orders being placed as the best guesses come into focus for market prices and for actual loads to be served.

Table 3-3 illustrates just one example of what the trader can face in price and volume uncertainty on a day-ahead basis. Purchases in each hour are shown as P and sales as S. Notice that there are multiple entries in each hour and the "net effect" is not easy to spot. For example, in hour ending 15 (3 pm) there is a 5 MW sale at $225 and a 50 MW purchase at $1,000. The net incremental transaction is $1,000. Once the forecast is known as to where this company will be in relation to that forward position, some amount of that transaction might be displaced. Now, let's look closer at hour 16 where there are three sell agreements. One apparently is the 5 MW block we looked at in the previous hour and two more 50 MW sell commitments are being made at $1,500 and $1,800 per MWh plus one 50 MW buy at $1,200 per MWh. Is the net incremental transaction at $1,200 or higher? This will depend upon where this energy company finds itself in relationship to the book of transactions.

Table 3-3
Example of Trader's Incremental Stack

All Activity In the Stack

Hr End	P/S	MW	$/MWh
15	S	5	$225
15	P	50	$1,000
16	S	50	$1,800
16	S	5	$225
16	S	50	$1,500
16	P	50	$1,200
17	S	5	$225
17	P	50	$1,800
18	S	5	$225
18	P	170	$1,800
19	S	5	$225
19	P	100	$450

Net Purchase Incremental Transaction

Hr End	MW	$/MWh
15	50	$1,000
16	50	$1,200
17	50	$1,800
18	170	$1,800
19	100	$450

Now let's return to the customer perspective and see how very upsetting this can be when customers simply read about (or worse yet, see on an ISO web site) market clearing prices at $1,800/MW in an hour while they are only offered a small portion of that price for their actions. The most equitable way to assure customer satisfaction is to create transparency and let the buyers and sellers compete for customer demand-trading capabilities.

Who Runs the Auction and How Do You Keep All Parties Whole?

One of the more interesting models for how demand trades can be aggregated and traded across and between organizations comes from the world of electric cooperatives. Here, the customer-facing organizations are the Electric Membership Cooperatives (EMCs) that are relatively small and generally serve relatively few large commercial and industrial customers. However, the EMCs read the meters and are generally the ones who would be responsible for customer-facing issues such as education, agreement negotiation, and customer service. These EMCs are served in turn by Generation and Transmission Cooperatives (G&Ts) who may, in turn, use wholesale trading organizations to interface to regional markets. Most G&Ts have portfolios of their own generation, possibly jointly owned generation, as well as supply contracts with other regional generating companies. These organizations do the supply planning and are responsible for most pricing decisions between member systems.

Each G&T-member EMC has a "rate schedule" upon which it purchases energy from its G&T for some or all of its requirements. Some EMCs are so large and capable that they will competitively shop for all or a portion of their power supply. Others, especially the smaller ones, will probably rely upon their G&T with full requirement agreements. The price signal their end-use customer sees is the bundle of their supply agreements plus whatever wires charges are deemed appropriate. And since these are EMCs, they are owned by their customers and any excess earnings above prudently retained earnings are returned to the members in the form of capital credits.

Ironically, this seemingly “one for all and all for one” environment is intrinsically quite contentious when it comes to demand response. The reason is that some EMCs learned that they could shift costs to other members by using demand response to reduce peak loads during those periods when the allocation of peak costs is determined. This can become especially contentious between member EMCs where some have industrial or agricultural customers that can significantly reduce loads while other EMCs have primarily residential customers where the load reductions were tougher (or more costly) to implement.

It is possible to compensate for this effect. Since each customer’s demand trades are settled individually, the amount their baselines have been decremented is well accounted for and the “normal peak load” can be inferred by adding back the demand trades before any cost allocations are made (if the trading days were in fact affecting this peak demand). So, any effects of customers performing in response to market prices in the region can be correctly factored into any allocations that would otherwise apportion costs differently.

Next, let’s move onto the question of how you allocate the benefits of mitigating price spikes in such complex multi-party situations. How can you manage transactions and send price signals along this complex value chain and keep everyone positive about contributing their piece of the value chain? There are obviously lots of ways to divvy up the benefits, but careful attention should be made to balance the rewards to reflect who is investing in an asset, taking market price risks, executing the trades, and being responsible for settling the transaction to all accounts in the value chain. Experience indicates it is best to start the offers to customers at something at least worth their attention – typically a minimum of $.15 per kWh ($150/MWh) as a benefit on their bill. That number can and should rise according to the displaced value, but how much should the trading organization get, the G&T, and the member EMC? The trading organization might normally get $1-$2 per MWh for initiating the trade, so that isn’t a deal killer, unless the trading organization is going to take title to the transaction and be responsible for all reconciliation and settlement expenses. The number we assume here is $1,000 per trade for a 100 MWh block (25 MW times a four-hour duration). There is generally a flat fee per trade to perform the baseline computations and produce a penny-accurate billing document.

The question then is how benefits should be split among the member EMCs whose customers participated and those who otherwise benefited indirectly. The split we have seen that seems to satisfy critics is to take the difference between the price of power in the wholesale market minus the price offered to the customer times the transaction volume and consider that the gross benefit for demand trading. The cost of executing the trade is subtracted from that, and the remaining net benefits are allocated to each with 50% of the net benefit flowing to the customer-facing EMC, 25% flowing to the G&T to recover the costs of implementing the aggregation efforts, and the other 25% flowing to all other members. This is shown as an example in Table 3-4.

One can suggest that different splits would be better and some level of horse trading is good to get organizations feeling comfortable with the concept. However, the key is that everyone "wins" and that the final price offer to the customer is automatically apportioned along the value chain to reflect benefit sharing as the demand-trading opportunities emerge.

Table 3-4
Cooperative Multi-Party Benefit Sharing Example

Wholesale Price $/MWh	100 MWh Wholesale Transaction $/Transaction	Trader Transaction Fees	Customer Price Offer $/MWh	Customer Savings $/Transaction	Net Savings In Transaction $/Transaction	EMC Benefit $/Transaction	G&T Benefit $/Transaction	Other EMC Benefit $/Transaction
$200	$20,000	$1,000	$150	$15,000	$4,000	$2,000	$1,000	$1,000
$300	$30,000	$1,000	$150	$15,000	$14,000	$7,000	$3,500	$3,500
$400	$40,000	$1,000	$200	$20,000	$19,000	$9,500	$4,750	$4,750
$500	$50,000	$1,000	$250	$25,000	$24,000	$12,000	$6,000	$6,000
$600	$60,000	$1,000	$300	$30,000	$29,000	$14,500	$7,250	$7,250
$700	$70,000	$1,000	$350	$35,000	$34,000	$17,000	$8,500	$8,500
$800	$80,000	$1,000	$400	$40,000	$39,000	$19,500	$9,750	$9,750
$900	$90,000	$1,000	$450	$45,000	$44,000	$22,000	$11,000	$11,000
$1,000	$100,000	$1,000	$500	$50,000	$49,000	$24,500	$12,250	$12,250

Avoiding Relational Confusion

By this point in our demonstration about how demand trading can cross organizational lines and fairly compensate all parties for its actuation, you might believe everyone is made happy. Not by a long shot. We have shown an example where all parties in the value chain know just what is happening as the price offers are made and the resulting demand trades are aggregated. We have also shown how all lost revenue compensation and cost shifting fears can be ameliorated. Such assurances are not always present in power markets at this time, but that situation can be improved as noted elsewhere in this document.

While competition for the customer's demand-trading capabilities is, in itself, a healthy force to discipline all market participants, it can potentially confuse and frustrate customers as they attempt to decide how to participate. And when customers get confused, they tend to do nothing. So, if the customer was on an agreement from their traditional regulated energy service provider, and some new market entrant makes an overture about another program concept, it is entirely possible that the customer will leave the old agreement and try to sign up for one they think will be superior. Then, as they discover this isn't to their liking for some reason (most notably that the promise of saving money was elusive), they are likely to abandon the concept completely (since they probably can't go back to the old regulated agreement). As a result, the free market's attempts to attract the customer to demand response can actually destroy part of the demand-response potential of the customers themselves.

There is no simple answer to this challenge. Choice is a good thing and this is simply one more of the challenges of offering customers choices. One might say that the best opportunity for energy companies here is to be an educator and an agent, thereby helping customers achieve the long-run best result for their demand-response capabilities, sharing in that through a performance-based partnership akin to an energy-services relationship.

4
ECONOMIC ANALYSIS METHODS FOR DEMAND TRADING

As the electricity market was being deregulated, the implicit assumption was that innovative retail solutions would naturally emerge as market players used them for competitive differentiation, and that a spectrum of price-risk differentiated products and services would soon result. Why hasn't that happened, and why is demand trading's role still lagging in today's electricity markets?

Part of the answer lies in the types of analysis approaches that have been used to assess the economics of demand trading. The commonly used traditional economic metrics often fail at accurately describing the situation at hand. They undervalue the protections offered by demand trades because they miss two key points:

- "Flaw of Averages:" The common metrics use averages to describe parameters that are not fairly represented by averages. Volatility impacts must be intrinsically characterized by the metric if volatility impacts the decision (and we know it does).
- Worst-case scenarios: Most of us are prone to underestimate the likelihood and consequences of adverse situations. Conservatism tends to be compromised when people try to build a business case for a concept. Perhaps this is due to the nature of capital allocations in modern companies. Everyone wants their projects to be approved, so many people try to sell their project as better than the next person's. Looking at the worst case, in addition to the expected case, for the alternatives under consideration can be quite telling.

This chapter starts with the traditional methods used to evaluate the economics of decisions and corporate well-being, discusses some issues surrounding risk and perceptions of risk, and then moves on to some new evaluation concepts, including one called *air worthiness*.[1] In the aircraft industry, airworthiness is defined as the ability to anticipate threatening situations and withstand unavoidable turbulence without destroying the airframe or harming those inside. By applying this concept to the power industry, we can develop new metrics of business safety and reliability that follow traditional business evaluation criteria (debt to equity, cash flow, etc.) but are different because they do not focus solely on average costs and benefits. As noted below, these new metrics consider the asymmetry of risks and rewards in the power industry and can help to analyze issues of economic survival. The metrics may be factored into code and therefore be

[1] Some of the terms used in this chapter come from the flight characteristics of aircraft, which parallel the risks and rewards of the energy industry in many ways. We all use the language of flight in business terms, so these should fit an intuitive framework for natural business discussion. You've heard people say: "That idea won't fly," or "Our business took off."

designed into market rules, and/or their conclusions may be socialized with everyone paying a small tax to help provide greater market stability.

It is instructive to look at the requirements of the Civil Aeronautics Board for a carrier that provides commercial air service. There are rules about the safety of the aircraft, but those are givens. The really tough requirements relate to the professional management of the operational and business risks of the business. Without such management in place, the carrier will not be permitted to operate.

Unfortunately, today's power markets lack air worthiness in many respects, and the rules to correct this situation are still a work in progress.

Traditional Economic Metrics

How is it that well-meaning professionals can be misled by seemingly solid economic predictions in today's power markets? The portrayal of risk in traditional financial metrics is normally handled well only when risks are relatively symmetric (i.e., the likelihood and consequences of things going well are reasonably similar to the likelihood and consequences of things going poorly) and averages reflect what can be expected. Since risks in today's power markets are generally asymmetric, with potentially large down-side consequences and lower up-side opportunities, the application of the traditional metrics often does not provide management with the proper information.

From another perspective, the traditional metrics have generally been used for incremental decision making, not to decide whether a business is a good or bad idea in the first place. Part of the reason for this perspective is the often-used assumption that any ongoing decision in business is incremental – not disruptive. In this bias lies the problem with energy risk decisions today. The impact of energy price volatility can be extremely disruptive.

The typical domain of the traditional economic metric is an index containing particular information about a series of receipts and disbursements representing an investment opportunity. The most common traditional economic metrics in corporate America at this time are:

- Present-worth amount
- Equivalent uniform annual worth
- Rate of return
- Payback period
- Life-cycle costing

These economic evaluation methods are all defined and discussed in Appendix A. This Appendix notes that the ways these methods as generally applied fail to properly alert management about the relative values and business risks of projects.

The Human Dynamics of Risk Portrayal

One cannot assume that the economic evaluation of an idea will always be performed using standard metrics and that management will simply select the best portfolio of projects using clinical attributes.

Engineers like to build things. Projects create excitement and activities that are naturally invigorating to an organization. However, given the resource limitations that all organizations face, management must be selective about their commitments to projects. For example, assume you operate a manufacturing facility that competes internally for production against other manufacturing facilities owned by the same company. You want to see investment for your facility over others because it is likely to ensure the well being of your plant.

All of this creates a natural temptation for project sponsors to portray their projects "in the best light." Given that most senior management will have neither the time nor the ability to check out all the assumptions, it is entirely possible that certain project attributes might be completely discounted, particularly if they do not enhance the project's chances of being funded. The most common one of these is risk itself. While it is generally assumed that projects will succeed, they often don't. Murphy's Law can result in a string of bad luck, but it is human nature to ignore this law.

Doing nothing can also put a company into an extremely risky position. Perhaps management should focus more on *position risk* and less on project risk – taking the position of simply doing nothing could be the riskiest thing to do. It is interesting to study industries that have long histories of operational risk and how they invest. The paper industry knows very well that they go through cycles of high and low prices. When they hit the cycle of high prices, they invest like crazy to get ready for the next cycle of low prices, knowing full well that only the strong survive that downturn. Their shareholders have come to understand this. For the power industry, there may be substantial value in the further education of retailers and other market players about the nature of risks.

Dealing with Infrequent Benefit Streams

The problem with many elements of demand trading today is that their economic value is akin to fire insurance, making them hard to justify as "investments." It is really loss prevention. If the perception is that losses are unlikely, the result is generally that little new demand response is acquired. This perception is ultimately tied to the value of capacity and the reserve margins in any regional market.

Additionally, uncertainty about the persistence of existing demand-trading arrangements looms large given the reformulations of the ISOs into RTOs or whatever new structures may emerge later. This regulatory risk confounds what was already a difficult situation.

With the various sources of risk that investors face, they do not know what the odds are going to be. They can only, at best, know what the odds have been. The investors can try to bound the situation by taking the worst, best, and average of what has happened in the past, but they cannot know for certain what will happen in the future. There is always a chance you can suffer a string of "bad luck" based on Murphy's Law. Explicitly bounding risk has value.

Let's look at this in the following example. Assume you were investing $1,000 and had two opportunities with 12-year lives that were equally risky from a technology deployment perspective. The only difference between the two investments is the payment stream. The first has a payment stream that is a flat $250 per year for the life of the investment. The traditional metrics for this would be a four-year simple payback and an internal rate of return of 23%.

Now, let's change the payment stream to every fourth year at $1,000 for the life of the investment. The simple payback period is still four years, and while the internal rate of return changes to 16.5%, the change is not dramatic. The biggest difference between the two options is the timing of the benefits. Given a choice, most would select the first option – the one with annual payments. Unfortunately, the situation of demand response is akin to the second option. There may not be any benefits in a given year, and the timing of significant benefits can be years apart.

That is why we have coined a new metric for these evaluations – the "mean time between benefits." Where this parameter is years, and the customer relationship agreements are yearly, the mismatch will inhibit demand trading. Where the relationship to the customer can be lengthened by either a performance contract or an installed asset, there is the possibility of annualizing the benefit to create a steadier value stream.

Said another way, creative energy companies will have greater chances of success in acquiring demand-response resources if they find ways to assure customers some level of payment for their demand trades on an annual basis. That will require a financial mechanism of making the market forward for demand trades. Energy efficiency, on the other hand, produces benefits all the time and, while the amount saved in any one year will depend upon the weather, etc., there is some level of confidence that there will be savings in any given year. While a third-party energy services agreement or some form of performance contracting might not be compelling in any one year, the 5- to 8-year term that these firms use provides a high level of confidence in the aggregate benefits to all parties.

Theoretically, someone can make any market forward by taking the long-term position. But, who can make the demand-trading market forward? One could argue that an ISO/RTO is best. Others might argue that the generation companies might hedge their bets by buying forward on the demand side – but how can they get to the customers? The relationship to most customers at the moment is through the regulated wires company, but wires companies may be precluded from working with customers in some areas of the country where competition for retail customers is underway.

This situation has happened in free markets before and commercial mechanisms have been used in the past. One of the most famous was in the shipping industry where boatloads of goods were occasionally lost at sea. In fact, the odds were that about one in ten ships never made their destination port. Their answer was that everyone who was shipping paid an extra 10% into a risk pool and, if their shipment was lost, the commercial value was paid. The risk underwriters were (and still are) wealthy individuals "willing to take the bet." The risk pool was an essential element in the creation of commercial shipping as a market, because unless there was a way to mitigate this risk, the market players wouldn't ship many goods. And, since the ship owners ran the pool with rules that were self-enforcing, it worked. After all, no one wanted to insure their goods at any other value than what they were worth.

The same approach could be undertaken in the energy business to make the market work. Someone would have to run the risk pool for the electric market participants in any regional market. The buyers out of that pool would be those who have loss exposures and they would pay in their dues in proportion to what they expect out. Later in this document we look at some of the issues of who might run such a pool and how.

The Human Dynamics of Risk Taking

Hopefully when people play the lottery they know that the probability of a significant win is exceedingly small. Why is it that anyone would play a game like this? It has to do with the contrast between the cost of playing versus the payoff size. It is a study in human dynamics and emotions about asymmetric rewards in which a small amount of money has a chance of becoming a much larger sum of money. The next time you talk to someone who has gone to the racetrack, ask them whether they bet the horse to win, place, or show. The best odds for being paid something back for the bet is to bet the favorite to show, but human nature will often prevail and the bet will be an all-or-nothing bet to "win."

The flip side of this risk profile is equally seductive – neglecting the small-probability and high-consequences risks. We all can fall prey to this one, especially as we start out in life after college. How much insurance do we carry, if any at all, on our possessions? What deductible should we have on our car insurance? How much life insurance do we carry? As in all of life, being able to withstand risks can save you money in the long run. We come to grips with that personally when we look at the deductible in car insurance. Can we personally withstand the negative cash drain should something go wrong? We will come back to that question as we look at the consequences of asymmetric risk in business and how a business case can be argued for demand response as insurance to mitigate that risk.

We really shouldn't be all that surprised by the difficulty energy companies have had selling demand response in their propositions. Selling insurance to those in need has never been easy, even where the benefits are clear and the customer would be imprudent to avoid it. Insurance agents have long learned that there are special times when we are all sensitized to risks and that they stand a much better chance of selling insurance to us at those times. That is why agents watch the newspapers for birth, graduation, and marriage announcements. They also know that the policies need to offer assurances and not dwell on the exceptions. "You are in good hands" is a phrase we have all heard. Energy traders need to think about this as they design structured agreements for customer demand trades. Liquidated damages, for example, is an expression that should be avoided.

If we are intrinsically so prone to **not** carry insurance, why is it that electronics retailers are now all on the bandwagon of offering electronics and appliance insurance as you try to leave with your purchase? The reason has to do with two factors:

- Retailers can make more on selling the insurance than they can on selling the device being insured, so they emphasize selling the insurance using incentives for their sales personnel.
- They make it easy to incrementally add the insurance at the point of sale. If they scared us about how much it would cost to repair the appliance and then gave us a list of independent insurance agents we could call to secure the insurance, very few would take the time to get it.

This type of insurance is described as having numerous potential benefits, all of which sound quite attractive, but few of which we will ever use. These benefits include free routine maintenance (which very few will use because almost no one keeps track of the time intervals that should be followed). Other benefits are the no-cost repairs and the "lemon" provision which covers a new unit if they can't repair the device to your satisfaction. However, the seller of this policy is fully aware that most problems will occur within the first few months of purchase and will be covered under the manufacturer's warranty anyway. It should also come as no surprise that when your unit finally does break, you will find that the service agreement expired a few months prior.

Telephone companies also sell this type of insurance to cover the phone wires in your home, and there are companies who will sell you insurance on all the major appliances (air conditioner, water heater, etc.) in your home.

Valuing the Unlikely High-Consequence Event

Why then is it that electricity price protections are not following the same course? Wouldn't an energy company use the same models noted above and sell protections to customers? Part of the problem today is that the customers have been able to force the various energy companies into "eating" these costs because the customer can lean on the POLR price as a price to beat. But another part of this is the lure that "it won't happen to me" and the belief that energy companies can "go naked" facing this risk because it seems unlikely and the consequential costs are so uncertain.

For example, just imagine that you are betting whether the toss of a coin is heads or tails. The odds are 50/50 and the chances of being wrong for 30 tosses of the coin are just too unlikely to be believed. This confidence in the "reversion to the mean," so prevalent in traditional trading strategies, is potentially disastrous when you have a run of bad luck.

People reading books about how to win at Las Vegas will often learn about doubling down on lost bets. If the odds are 50/50 and you lose, the concept is to double the next bet. If you win, you are ahead, and if you lose again, you double the bet again. After all, you can't have a run of continued losses, can you? Las Vegas is aware of this strategy and will limit the number of times you can double down. When you do play this game and hit that limit, your losses will be so great that all the other previous winnings will seem rather small. This is a very real illustration of the unwillingness of people to face the fundamentals and to think they can beat the system.

Another part of the problem is that in real life, as opposed to the situation in Las Vegas, the underlying actuarial statistics are less certain, or at a minimum, certainly less well known. This wasn't lost on energy trading companies in recent years, some of which had large cadres of doctorate-level statisticians and meteorologists on staff to take the "opinion" out of their forward positions and premise their positions, to the maximum extent possible, on reasonable math. But, even then, the mathematical process is largely backward looking. It would have been unthinkable for someone in 1997 to suggest that hourly spot market prices could ever reach $7,500 per MWh in 1999 and $8,500 per MWh in 2000.

One might argue that energy companies that lack insurance by "going naked" present a public disservice – analogous to the situation of someone lacking car insurance – in that it puts the general public at an increased risk of high prices and power shortages.

Risk Valuation: Bringing Risk to Light

In any commercial agreement, there is almost always a clause covering the limitation of liabilities. Lawyers involved in negotiations over this clause may seem to be "inventing" risks, and the uninitiated can see them as unnecessarily raising fees and starting arguments where there weren't any to begin with. While this is possible, most lawyers are simply trying to bring risk situations into reasonable negotiation about their resolution before things go wrong.

Energy companies in open, competitive markets should also have these dialogues with those they serve, but seldom do. Many such companies are more concerned about signing up customers than bringing market risks to light and potentially complicating negotiations.

The Dangers of Turbulence for Aircraft and Markets

Whether a Chinese curse or not, we are living in interesting times. One could assuredly call them turbulent – a nice word, but what does that mean? Following the aircraft paradigm, turbulence is something other than the mainstream air movement. We have probably all encountered the "bumpy ride" on an aircraft, and most of the time it was simply a nuisance because it forced the captain to suspend service in the cabin.

At some level of turbulence, the ride becomes "unsettling" to those who had never experienced it before and such passengers often look around to see whether it is bothering anyone else. If most people are continuing their normal behavior, the unseasoned traveler is likely to grin and bear it. If, on the other hand, people around them are screaming, it is likely that the unseasoned traveler will follow suit. The captain and crew, on the other hand, may know full well that nothing is seriously wrong and the plane will hold together. They are experienced and have weathered these situations countless times in the past. However, it is imperative for them to communicate that to the passengers who are certainly uncomfortable.

As the scale of turbulence increases, there will come a point where it really is "stressing" the aircraft and ultimately depleting its useful life. Over time, the accumulated stresses will eventually take their toll, resulting in cracks and component failures. Routine inspection is the only way to know about this and skimping on maintenance can result in disaster.

It is one thing to damage the aircraft, land without fanfare, and simply replace or repair the failed part. It is quite another to have the failure result in a crash, and worse yet, fatalities. The costs of a crash include not only the loss of the aircraft and lives – they also affect the image of the carrier, the aircraft fabricator, and the entire industry. At a certain point, if failures become perceived as unacceptable, the regulators will ground the fleet (e.g., the L10-11 and the Concorde). That is why the Civil Aeronautics Board (CAB) will not permit a carrier to provide air service simply because they have safe airplanes. As noted earlier, carriers have to prove to the CAB that they have the full bevy of systems, procedures, and financial underpinnings. Perhaps this concept of "air worthiness" should have been considered more carefully when decisions were made to open some regional electricity markets.

In the airline industry, when the costs of inspections were simply cost-recovered as "prudent" in the old regulated business model, these inspections were assured of being performed. As the industry became more competitive, the likelihood of these inspections being rigorously performed was reduced. That is why we generally force third-party inspections in these areas of public trust. However, this can only occur where we know that risks exist. If we have not seen certain types of risks before, we generally do not know how to inspect for their presence, and probably have no idea of the consequences. This seems to reflect where we are in the power industry at the moment.

Unfortunately, the risks facing the power industry go beyond the consequences of not achieving quarterly goals. They can, at some level of market turbulence, cause businesses and system reliability to crash and burn. Constructing the proper business model in a competitive market, therefore, needs to not just focus on minimizing average costs or increasing average profits, but also on the issue of survivability.

Survivability and Staying Power

A simple, but powerful approach that can be applied to analyzing a business strategy in the power industry involves the question of surviving the worst-case scenario in any one year. This is not to imply that this business case is the likely outcome, but this question is one that sober business professionals should ask. Can they survive and gracefully exit a worst-case scenario?

Worst cases are often the combination of many variables all going wrong at the same time, in total harmony with Murphy's Law. The key to thinking through these situations is to ask "what if" and to avoid the phrase "I think." At some point, the assumptions may seem absurdly skeptical and conservative, but the contingency plans are essential to a reasoned relationship with management and the investment community.

One should consider survivability in the context of non-independent risks. For example, an insurance company writing storm coverage for the entire U.S. has a much better risk profile than one just covering the State of Florida. If a hurricane hits Florida, the firm underwriting Florida alone could be wiped out by claims if it had not hedged that risk through reinsurance. Interestingly, it is likely that the cheapest insurance you could buy would be from a firm like this, but it is also possible that you would not have your claims paid when the hurricane hits. The same flawed logic is occurring in today's energy markets.

Another approach to consider is the ability to live through a "run of bad luck." External factors can destroy the best of forecasts and even the best sales team can fail to get customer commitments when customers are distracted by other agendas over which you may have no control. One can talk about the "staying power" of a company and the "war chest" it would need. The war chest analogy may be apt because it implies that the situation is as potentially disastrous as a war. You really do not want to go into any battle where both sides are about equal – that would be extremely costly to both sides, and not worth the encounter.

The acid test of the combination of these factors is a conservative and realistically pessimistic view of business conditions on the cumulative cash flow requirements. Can the company withstand that occurrence and does it have that cash in the war chest ready to withstand that outcome?

Asymmetric Risk

Asymmetric risk is an important concept in today's power markets. Compare the two probability distributions shown in Figure 4-1 from a cumulative distribution perspective. Notice the symmetry of the normal distribution and the right-hand tail of the log-normal distribution (which tends to be how commodity prices behave).

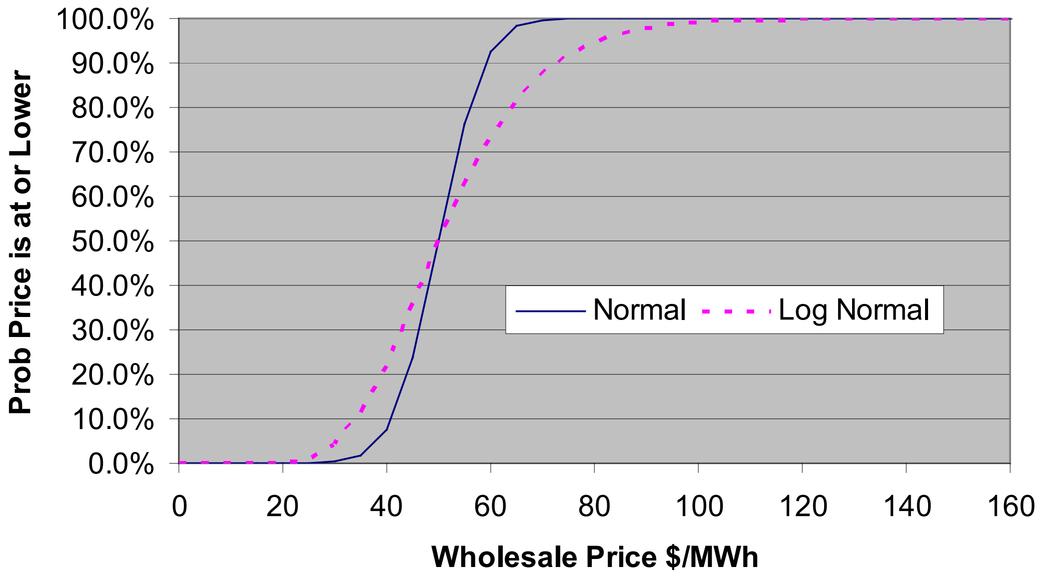

Figure 4-1
Comparison of Normal and Log-Normal Price Distributions

The result is that while both distributions have the same probability that prices will be at or below the median price of $50/MWh as shown on this graph, the log-normal distribution has a significantly higher tendency to produce high prices. For example, the cumulative chance that the price will be $60 or higher is 10% if the distribution is normally distributed and about 30% under the log-normal assumption. While the highest prices might be tolerable in the normal distribution, it could be devastating in the log-normal case.

Said very simply, when things go wrong, they can go very wrong. In fact, volume risk compounds the impact of this price risk and the combined effects can be brutal. For example, Figure 4-2 shows a profitability risk evaluation for a potential energy company serving a particular customer. Three sets of business scenarios are shown: a normal year (where the historical average weather is assumed), a milder-than-normal year, and a severe-weather year.

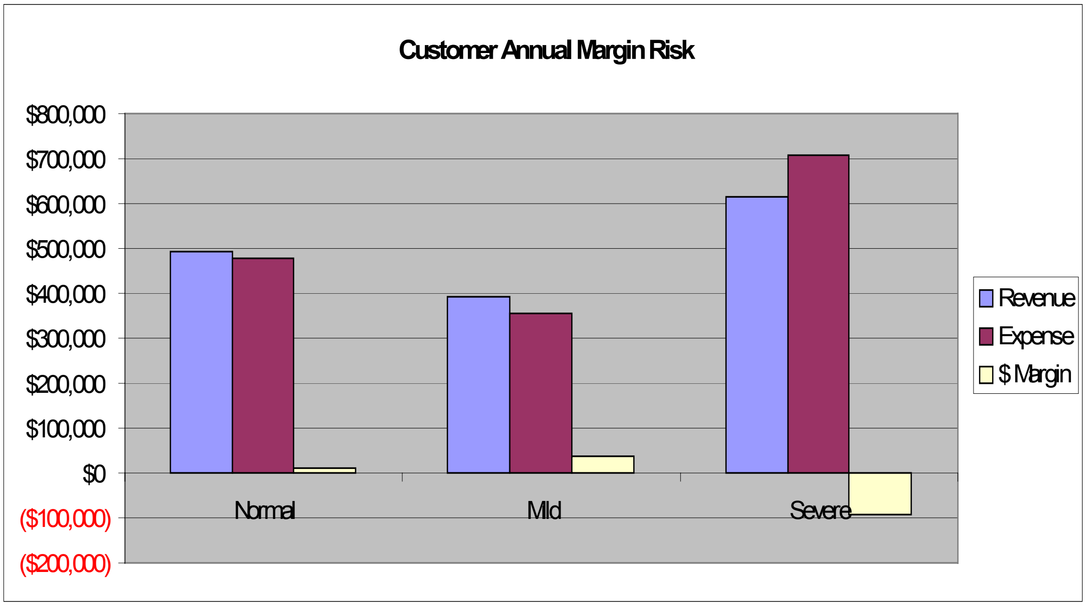

Figure 4-2
Retail Margin Variations with Weather

Notice that the margin in a normal year is small in comparison to the revenue and the purchased power expense. Better margins are achieved when the weather is mild, but the negative margins that result from severe weather are devastating. While the severe-weather year might only occur every five or six years, the unprepared energy company may not survive this type of year. By the way, these calculations do not factor in the business overheads of acquiring the customers, customer service, metering and billing, and any collections. These are simply the commodity costs and margins.

Required Staying Power

As noted earlier, another key attribute worth considering is a company's staying power – the cumulative cash required to stay alive through a run of bad luck. Where demand trading can mitigate that total cumulative cash requirement, it would represent an option an energy company should consider. However, the probabilistic nature of the analysis can frustrate intuition since most of us cannot conceive of a string of bad luck until it confronts us. For example, take the unhedged annual margin data from Figure 4-2 and assume that the severe weather only occurs once every six years, that the milder-than-normal weather does the same, and that the weather is average four out of six years. We could have the expected performance over one six-year period that is shown in Table 4-1. In this table, the hedged-margin case mitigates the losses during a severe-weather year.

Table 4-1
Deterministic Comparison of Retail Propositions with Weather

String of Events	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	
Weather Type	Normal	Milder	Normal	Severe	Normal	Normal	Average
Unhedged Margin	$12,722	$36,317	$12,722	($94,251)	$12,722	$12,722	($1,175)
Hedged Margin	$12,722	$36,317	$12,722	($47,125)	$12,722	$12,722	$6,680

Even though the average of the six years for the unhedged case shows a seemingly small negative margin, one's intuitive impression is likely to be that a bad year can be survived – on average. If we now compare that to hedging the adverse condition that only occurs once in six years, you can see the clear value of the improvement on average. It should be kept in mind, however, that there are probably lots of ways the energy company could hedge, and any costs to do so would have to answer to the apparent lack of value in most years when the value of the hedge was zero.

Let's now look at a probabilistic analysis of this situation–one that can provide the decision maker with a better understanding of the risk characteristics. We modeled this situation in a retail setting by comparing the maximum cumulative cash flow requirements for the retail business in both hedged and unhedged positions. The results are shown in Figure 4-3 as cash required to run this business for five consecutive years. As one might expect, the results are quite variable and can only best be illustrated using probability distributions for these cumulative cash flows.

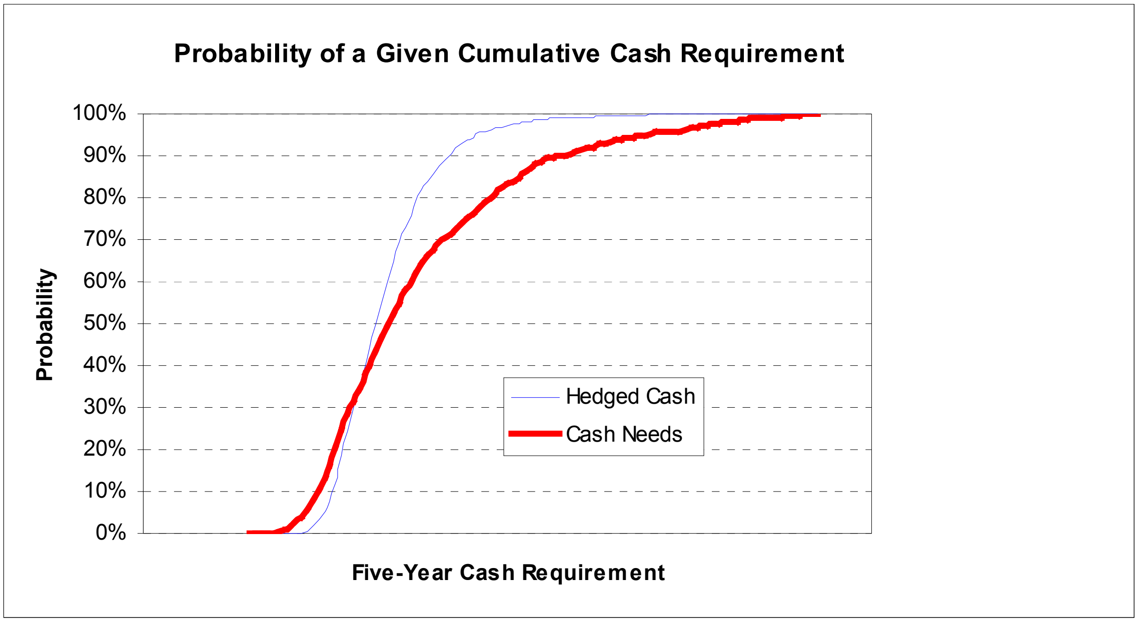

Figure 4-3
Cumulative Cash Flow Requirement Comparison

While the specific assumptions for this can always be argued, notice the general result is exactly what you would intuitively expect. The hedged case (the thinner line labeled "Hedged Cash" in Figure 4-3) mitigates the worst-case cash requirements in comparison to the unhedged case (the thicker line labeled "Cash Needs"). Notice that it does not affect the average performance as much as it prevents the extreme cash needs. Said another way, hedging provides significant staying power.

The only remaining step for the retailer is to consider whether the acquisition and operational costs of securing this hedge are in balance with this improved risk profile. In addition, as will be mentioned in a later chapter, consideration should also be given to the first-mover advantages. For example, how much worse off would you be if you chose not to acquire demand trades and a competitor had "skimmed the demand-response cream" from the regional market? In many ways, the business case for the next retailer is worse because the cost of acquisition will be higher and the resource quality is likely to be lower. Given the limited nature of the total demand-trading resource in most regions of the country, the first mover advantage can be substantial.

This is not really that unusual a situation as new markets develop. The successful first mover usually dominates the market (if they can survive the startup). Given that demand trades limit the downside risks during that startup, one should therefore consider them as useful insurance, not discretionary investments.

Getting Control of Electric System Planning and Operations

Sometimes we fool ourselves and think things are under control, when in reality, they are not. For example, when we hear that electric supplies are adequate and that price spikes are unlikely, we can become lulled into comfort and complacency. In many ways, these quiet periods are often nothing more than the lull before the storm. There once was a time when the electric utilities would look out into the future and see a need to build generation.

What did they build? Once nuclear was ruled out, it simply became a choice between coal and natural gas in most cases. Despite the lengthy lead time and the greater capital intensity of coal-fired plants, they were often preferred because coal was relatively inexpensive, coal prices were not volatile, and you could store substantial quantities of it. Environmental challenges were more difficult, and getting more so, but the direction was still to build coal plants.

The oil shocks of 1973 and 1978 and Three Mile Island in 1979 caused utility executives to reduce their plant construction schedules and independents entered the fray. The independents largely built combined-cycle, natural gas-fired power plants with shorter construction times and were able to pass along the fuel volatility or bought futures agreements or wellhead gas contracts. Now the planning horizon was more manageable. The counterparties to these plants were traditional customer-serving energy companies who would now pass along these costs through their fuel clause agreements or in their normal cost-of-service accounting.

The shortened timeline for construction, plus the smaller size of these plants (typically in 200-500 MW increments rather than 1,000 MW and larger blocks), improved the feedback and control loop in the industry and all seemed well. In addition, as the pressure to keep rates down and the enthusiasm about building waned, these energy companies built interconnections to each other in order to improve reliability.

The process also included the serious consideration of energy efficiency and load management as alternatives to construction and as a way to reduce the rise in electric loads over time. The implementation of these alternatives was based on integrated resource planning (IRP) and least-cost planning approaches. Figure 4-4 captures this process over time and shows the conceptual transition from large central stations toward independent power producers (IPPs) along the way.

The forecasted demand is shown moving upward over time but at a varying rate (due to the natural fits and starts in the economy). The old world planning horizon would consider the plant-construction time and then estimate the demand plus reserves (normally assumed at 20% in the old days) and construct the theoretical supply needed. The physical capacity in the old days consisted of lumpy 1,000 MW or larger generators. As the size of those generators and the planning horizon shortened, the firm supply requirement and the theoretical supply requirement were both adjusted to more closely match actual demand. The result was, as the graphic indicates, a less “lumpy” and more accurate construction plan with the introduction of the IPP.

The assumption today for new construction (in most areas of the U.S., and with the notable exception of Florida) is that the free market will construct new generation based upon market price signals. At this time, Florida is not planning to see a deregulated market in the near term, but we should follow what will happen there (with a regulatory control strategy) with what has already happened in much of the rest of the U.S. where market forces are being asked to provide control.

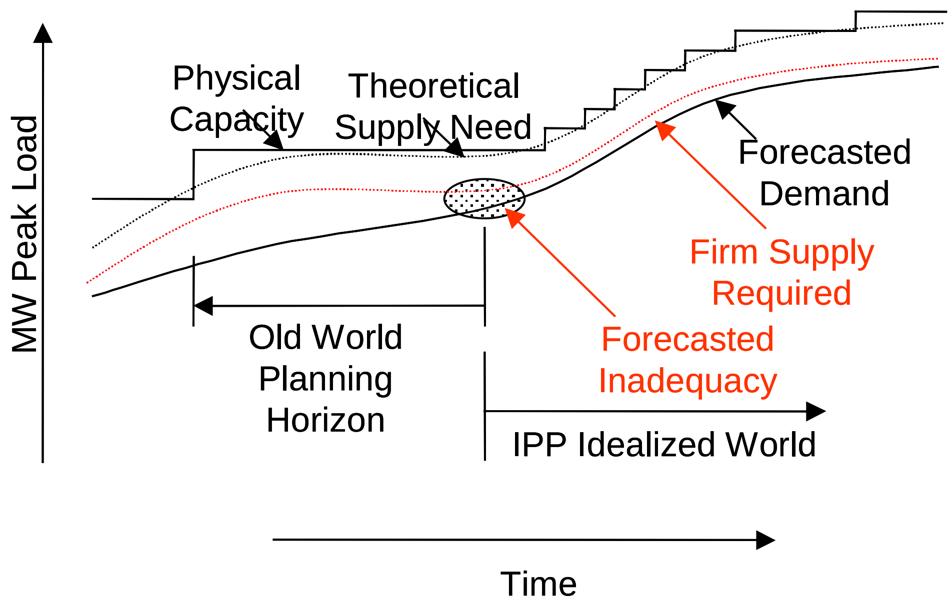

Figure 4-4
The Traditional Supply-Side Planning Model

Anyone who has watched wholesale price signals from the period of 1997-2002 will certainly attest to their volatility and to the reality that supply or delivery system inadequacy results in high clearing prices. Where these inadequacies are simply transient anomalies due to the weather or unusual physical failures in the delivery system, one might dismiss them as just that. However, a key question this section will discuss is whether we can really use market price signals to trigger the power-plant or delivery-system construction processes.

All control systems require an error signal to trigger actions. We are assuming a high wholesale price is a required indicator for new construction. If there were a liquid forward wholesale market (as there is in the natural gas business), the forward price curve would signal the capital markets. When those prices are anticipated or seen, the free market acts using the "invisible hand" so eloquently described by Adam Smith.

But let's take a look at what that implies. Today, the lack of forward business model certainty is causing the operational parties in the electricity value chain to wait for clear signals. Now we are waiting longer to act and often the majority of the market will only respond when the error signal is "clear and certain." We have now moved from a leading control signal to a lagging control signal. That automatically means we are going to face a boom-bust cycle for generation.

Now we wait for the error signal to rise, for it to be perceived, and for the market to act on it. Given the difficulty new generation faces in everything from locating an acceptable site, to gaining regulatory approvals and interconnection agreements, this error signal can be expected to persist for some time before being corrected by actual generation construction. By the time that happens, it is also quite likely that the error signal will have reached such a high level that it sparks a crisis and causes widespread criticism of the price signal itself. The history of the past few years is replete with examples of this outcome.

The idea of using price signals as a control introduces a potentially disruptive lag in the planning cycle that can result in unacceptable volatility. We seem to know that now, but the real question is what can be done to improve the situation? How can we build a better control system that "anticipates" price movements and invests in resources to prevent price spikes, rather than races to apply Band-Aids and tourniquets to market disasters? If the NYMEX contracts for electricity had worked as they do in natural gas and oil markets, we could point to them and use them for long-run price signals. However, the NYMEX contracts for electricity failed all across the U.S. and this failure suggests an intrinsic problem.

Free-market risk capital relies on forward market signals. In a free market, simply insisting on reserves doesn't necessarily result in their construction. Some might argue that the best answer is to return to a greater level of regulation, requiring the traditional regulated wires companies to answer to their Public Service Commissions on minimum reserves and IRP in prudency audits. However, this would truly defeat the movement toward retail competition as we currently understand it. Another answer is to insist that energy companies provide adequate reserves to meet their obligations, but today's POLR tariffs do not reflect these capacity costs, and forcing the energy company to do this will diminish what is already a fragile retail proposition. Another answer being considered is to force the POLR provider to move customers to time-differentiated pricing mechanisms such as Real-Time Pricing (RTP) so that, if nothing else, the customer is now made aware that the prices are not the same all year long. In this way, customers can save money by changing load shape.

It can be constructive to illustrate how the process can be controlled in other open, competitive markets, and to examine what is missing from today's electricity markets. This analysis can suggest how those control mechanisms might be introduced into the power industry and provide insights into the implications of such mechanisms for liquidity in demand trading. The good news here is that electricity is too important a resource to leave unaddressed and the industry has an enormous talent pool working to resolve these issues.

Planning to Avoid Price Turbulence

Many pilots humorously describe flying an airplane as thousands of hours of boredom interrupted by a few moments of sheer terror. The same might be said of electricity markets. Extreme price turbulence is unlikely, but potentially devastating. How can we anticipate it, fly around it, and when all else fails and there are no alternatives, fly through it safely?

The obvious preference is to avoid it altogether. In demand-response planning, this "best-of-all-worlds" solution is where there is enough demand response in the portfolio that there will be no price spikes. But wait a minute – if we are paying customers to avoid price spikes, and there aren't any, where are the economics of doing that? How can that be a free-market solution, or is it? It turns out that it is, but it takes a little more thought and control theory to get there.

It stands to reason we must have "avoided" an even higher cost by doing this. This baits the question, how much demand response is enough? Is it possible that you could develop too much demand response and potentially kill the economics for it through an excess of demand response in the supply-demand balance? Of course that is possible, and has always been recognized in traditional program designs for curtailment and interruptible programs. It is only to be expected that a free-market retailer would also ask the same question about program size.

These are all very important questions to consider. Just as in the amount of added insulation you could consider adding to your attic, there is a law of diminishing returns. One might easily argue the intellectual premise that the amount of insulation anyone adds to their home should be the acceptable incremental payback of the added insulation above code requirements. Each inch of insulation has to answer to the incremental savings. That is, it is up to the homeowner to decide – or is it – or should it be? What about the possibility that the collective actions of homeowners, acting on their own economics, might not choose the best options for society as a whole? Under these conditions, might incentives be created for the insulation of homes based upon the societal value? Or, might the standards just be tightened on all home construction to be sure customers do the right thing?

Not long ago, energy efficiency was viewed as the first option many electric utilities had to pursue and these utilities had to pay customers not to use their product. At the time, some customers complained that these incentives would raise prices. But at many Public Utility Commissions, it was decided that in the long run, overall costs would be lower than what would have otherwise been the costs with additional generation. Some utilities also objected until shareholder incentives were allowed for energy-efficiency programs. As the CEO of a utility said at the time, "the rat has to smell the cheese!"

While some might suggest a move back towards central planning and the socialization of benefits, it is likely that greater benefits will be obtained through movement toward open market models. It is interesting to note, however, that the spot market does socialize certain costs whether we want it to or not. The Dutch auction process of giving all bidders the final highest cleared price not only brings forth the most cost-effective resources, it also imposes the final cleared price upon all of those buying in the spot market. When the spot market goes into constraint and prices move higher, the result can easily be shown that avoiding that last MW can not only reduce the cleared price, but can also produce significant collateral benefits to all other participants buying in that market. Regulatory bodies have picked up on this and are considering demand-response benefits using this model.

This raises the related question that some have always asked: does a free market reduce prices over a regulated market? The answer to this is that free markets reduce prices over regulated markets *when supply exceeds demand*. When supply is inadequate, free markets operate at higher prices than regulated markets. These high price signals during inadequate supply periods bring about an eventual increase in supply, and in the long run produce lower prices than regulation. (Note that the blocking of high wholesale price signals can derail this mechanism.) We know that another advantage of free markets is the innovation it naturally produces. For example, not too long ago, telephones were all black.

In many ways, we are all living through one of the most dramatic real-life experiments in free market economics. Many suggest that the free market should be allowed to work, but the implementation in certain areas has caused a difficult mix of economics, regulation, and politics.

Control System Approaches for Power Markets

An important initial point is that it is not easy to apply control theory in systems that require customer consideration of energy issues. The supply side of this business is physically tied to energy issues and must pay attention to them because that is its business. The demand side is predominantly composed of people who are only occasionally sensitive to energy issues. Plus they have emotional insensitivities, bias, and inertia about their energy attitudes that can confound the economic theories that would indicate customers do things because they can save money. While the demand side can be made more physical through the introduction of technology, it will still generally require other human agents to interact with the markets.

The first point in any control system is that price has to be permitted to signal the market. Policies that limit high prices inhibit the economic principles of open markets and while high prices can be painful, they are generally required to bring about the development of the required resources that lead to the promised land of lower prices. There are other ways to get past this impasse, but they will require some mechanism to level out the near-term and longer-term price signals.

The simplest control is an “on-off” system that brings in a fixed response to an error signal (such as deviations above or below a temperature setpoint on your thermostat). Once that response corrects the error, the fixed response is terminated and the system waits for the error signal to build again.

While this may be an adequate method for temperature control in many conditioned spaces, it fails miserably in more-complex processes. This is why most real-world systems are not simple on-off controls. They can instead operate on a proportional response where larger errors precipitate larger corrections. This approach is not entirely adequate, however. There is reason to include other factors into the control system that reflect the history of the error signal as well as the way that error signal is changing. For example, where the temperature in the room is rising, a simple proportional signal will “lag” and literally never catch up until the rise terminates, at which time it will over-correct. In addition, even where the control point is constant, proportional control diminishes the response as the error decreases and “you never quite get there.”

This simple example indicates why controls move beyond simple proportional models to include the history of the error signal (integral effects) as well as the change in the error signal over time (differential) to truly respond to dynamic conditions. These systems are referred to as PID (proportional, integral, and differential) controls and are the standard in true-to-life complex control systems. We have several proven models of dispatching resources using these control methods in the electrical systems of today. ISO/RTO operations "filter out" the ramping up and down signals needed to balance the hourly markets in a given day from the instantaneous variations and feed these signals to appropriate counterparties in the markets. In addition, this balance was planned well ahead of the day itself to be sure there were adequate resources available.

Unfortunately, these control systems require an error signal before something happens. If the market has to wait for high prices to kick in demand trades, the control signal may lag rather than lead the situation. For example, if the regional markets do not have day-ahead market price discovery, the customer is forced into a day-of price signal, and possibly one in near real time. When the price signal is offered, but customers cannot act upon it, perhaps we shouldn't even think of it as a control signal at all.

It is interesting to note that Public Utility Commissions used to set reserve requirements at levels that prevented price spikes and held regional energy companies responsible for their obligation to serve. That forward planning model seems to have been replaced in many areas of the country with a model that implies regional price signals will precipitate the right level of generation and in the proper locations. Since that price signal to the capital markets is not coming from forward bilateral agreements (which might have signaled the correct result), but rather from some assumptions about regional market prices, the obvious happened. Where prices were forecast to be soft, inadequate generation was built.

The idea of a simple reserve margin is an inadequate control model for an open, competitive market. We really have three "error signals" in the power system: bilateral forward, day-ahead and day-of hourly (essentially ramping agreements), and ancillary services. One can imagine them as three separate control systems looking at different time scales. Forward markets are the essential and primary control signal. These markets should dominate the overall control and have a price signal that triggers the financing, design, and construction of new capacity. Once the forward markets function properly, the other markets have a chance. If the forward markets fail, the others run into difficulty.

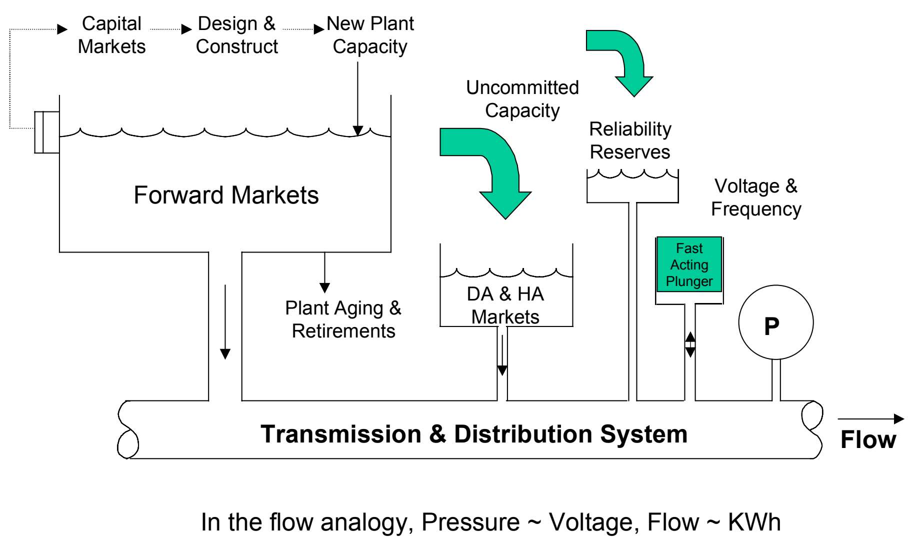

Figure 4-5
Interaction of Basic Electricity Market Signals

The diagram in Figure 4-5 shows a "fluid analogy" approach to depicting these markets and how they interact. While this analogy has its limits, it is a common-sense physical model that most people intuitively understand. The big pipe is the "grid" which has to be maintained at the proper pressure (equivalent to voltage) and accommodate the required flow (kWh). The large tank represents the physical generation capacity. When its level is sensed to be inadequate, capital can be invested to bring on more capacity. When it is adequate, uncommitted capacity is made available in the shorter term (day-ahead and hour-ahead) spot markets, along with the provision of ancillary services. The relative sizes of these markets are characterized by the sizes of the tanks as shown.

If the forward market fails, there will be inadequate capacity to meet day-ahead, hour-ahead, and reliability requirements. This is the basic control signaling the capital markets. Forward markets are the foundation of all planning and today's electricity markets are not sending reliable long-term signals at the moment – that specific block in the control system is broken. There is too little confidence that capital investment will provide adequate returns; the consequential undercapitalization could have severe impacts.

Lack of investment confidence has been experienced in other industries as well, and there are several ways that control systems can be stabilized and made to work in such situations. The list includes government guarantees (e.g., crop subsidies), preferential tax treatment (e.g., investment tax credits for renewable energy resources), low-interest loans (e.g., rural electrification), etc. These are the "gain" (or amplification, if you prefer) components in a control system in that they can increase the response to any given stimulus or error. However, there must be an error signal the markets can trust for commercial confidence.

There are certainly plenty of examples where price or value-based error signals would be socially undesirable. In these situations it is common to see "criteria" response mechanisms used that simply mandate all market participants to act. Examples of this include energy codes for new building construction and reserve requirements for retail market participants. When all market players are subject to these criteria, they work, but when only a fraction of the players are subject (such energy service providers alone) and they can slam customers back to those who do not face these criteria, they don't work.

PID Control System Applied to Demand Trading

The application of a general proportional, integral, differential (PID) model requires two foundational enabling elements: a criteria (or setpoint) against which to measure the error signal (i.e., the degree and direction for change or control), and the ability to provide an error response in adequate time to produce a beneficial result (as measured against that criteria). For example, you have to know how well you are doing to know you are on track, and have reached the desired result. It doesn't do much good (other than to assign costs) to measure the opportunity as it has passed us by, except possibly to validate the need for a better measurement system and to accumulate the potential value lost over time.

The ultimate goal of deregulation is to use a market-based valuation as the signal for financial gain, and that valuation – price – is, at the moment, extremely volatile and temporarily stuck at an unsustainably low index. Let's assume a long-run valuation (and a buyer at that valuation) can be structured from a market agent willing to take the variability out of forward value volatility. We investigate one specific model in Chapter 6, which involves regional demand-response reserve banks, but there are several hypothetical models. In all cases, these alternative models would have to resort to the same financial mechanism: converting the infrequent, unpredictable, and highly volatile value for demand response into a predictable value stream for interested market participants. In any event, let's assume there is a buyer, so there is a long-run valuation and that valuation is attractive for any alternative that would cost less than a generator.

This valuation is not a single-point comparison but actually a continuum from baseload to peaking power. This can be best summarized in Figure 4-6, which illustrates power costs versus the number of hours in a year during which two different types of power plants are used. It turns out that this plot is typical of all insurance premium models: you pay more for infrequent, high consequence protection. The illustration here is simply an example and does not include any local transmission, distribution, or environmental factors. The actual valuation for any region of the country depends upon the capital and amortization costs along with fuel and operating and maintenance, but the general shape will always be the same. Notice that short hours of annual use favor the combustion turbine due to its lower capital costs.

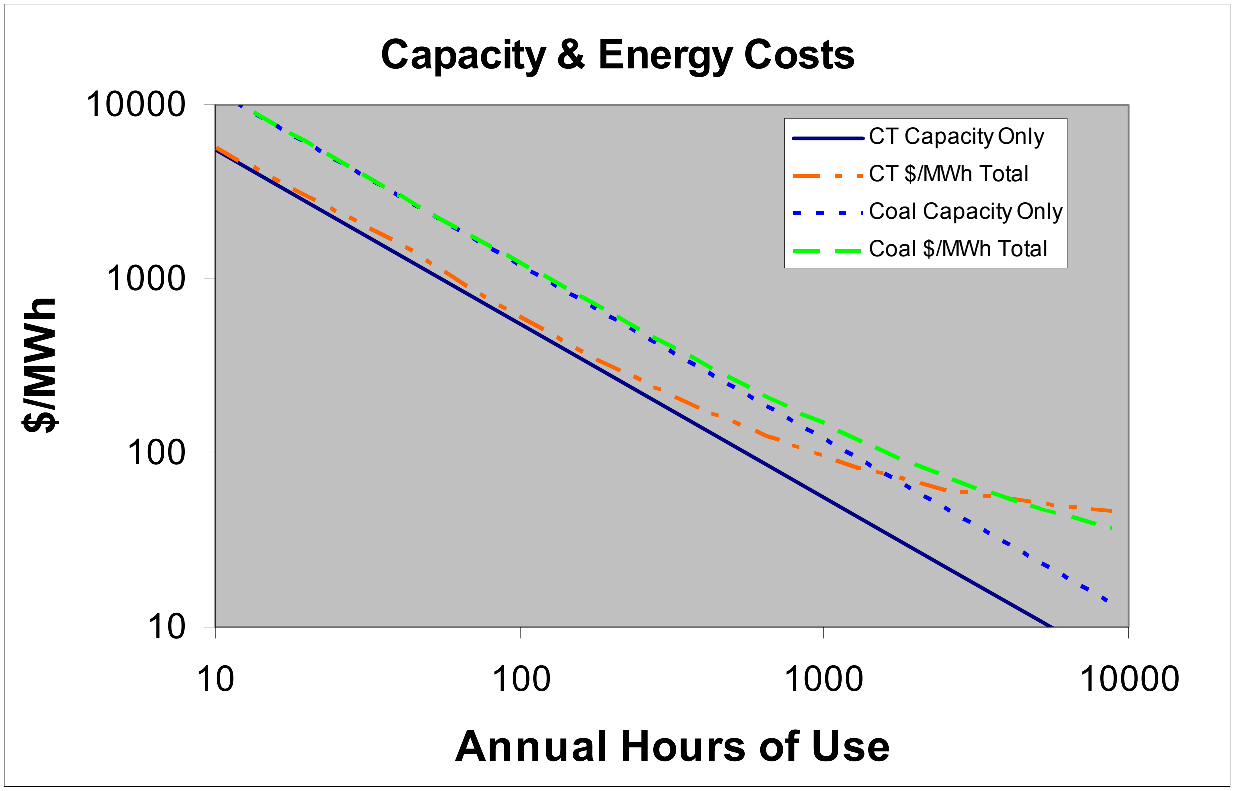

Figure 4-6
Power Costs vs. Annual Hours of Use

Both the supply side and the demand side view this graph as measuring the "opportunity" to build, and have to make the investment decision about whether to build or buy. Without a control system, the supply side would look at this chart and possibly overbuild and thereby destroy the value proposition itself. That is why the regulated electricity industry had a certificate of need underlying each and every power plant built so that an appropriate construction schedule would result. The assumption today is that wholesale electricity markets themselves produce the appropriate control signal – something clearly refuted in earlier discussions in this document.

So, what would the elements of a PID system look like if there were a long-run fundamental value all could see and use to plan appropriately? As a simple case, let's start with customer energy efficiency, which is an alternative to baseload generation. Customers (or their agents) would pursue energy-efficiency and fuel-switching opportunities to the level it made economic sense to do so, with full awareness of the investment risks. If the price forecast had reasonable certainty for a period of five or more years, the customer would probably have an excellent idea about what level of investment made sense over that period and would investigate/implement opportunities.

The individual feedback elements in PID are proportional, integral, and differential. The proportional elements are those where the "error" or "opportunity for improvement" is communicated in a linear proportionality. As we illustrate the concept, we will discuss two control system models: the customer's and the market counterparty's. Customers, by way of swap agreements with the suppliers, no longer see a time-differentiated opportunity. They instead see a near-constant valuation. So, the customer's perception of opportunity (error signal) emanates from the net incremental savings over implementation costs at any point in time they have through energy efficiency. Due to the nonlinear nature of the quantitative (e.g., cost vs. benefit) and qualitative (e.g., nuisance) impacts of such changes on business operations or lifestyle, the attractiveness of making changes in operations to achieve those improvements declines eventually to the point that the customer simply ignores further improvements.

As a result, and as is characteristic in all proportional control systems, the error signal diminishes as the customer implements the opportunities and the system never quite gets to the setpoint desired. Therefore, from the customer's perspective, the proportional response elements in this control system would be the financial performance of the energy-efficiency measures themselves as perceived and brought into convincing monetary valuation. The engineering expertise and complexity involved probably require that a third-party agent act on behalf of most customers; performance contracting is a proven mechanism to establish and retain such partnerships. To the extent that these opportunities were highly attractive, customers could pursue them individually.

Since proportional correction alone "never quite gets there" it is helpful to sensitize the customer to the small residual errors remaining after proportional response measures are taken. This is accomplished by integrating the error signal over time, keeping track of the additive effects of these potential saving opportunities over time. Tracking the integral thereby eventually compels those customers with many small residual opportunities to pursue them either themselves or use agents to arbitrage the benefits of their implementation. In the commonly used metaphor, if the integration signal was missing in the control system, like a screen door, the customer would never quite close in on the full extent of the energy-efficiency opportunities.

The purpose of a differential controller is to alert the system that something unusual has happened and thereby needs special attention. Said very simply, the control system has noticed that something is different at the moment. Since in our illustration so far, the customer has been insulated from the time variations in the wholesale markets, an example would be when the "opportunity for savings" signal picks up something unusual. The most common customer situations where this differential signal (which arrives from the business market in which the customer is engaged, not the electricity market) triggers "time sensitive" opportunity situations are where a customer has already decided that they need to renovate a space (which would make a lighting retrofit both more cost-effective and politically palatable to occupants) or replace major equipment. A differential controller senses a unique opportunity for correction by differentiating the time-dependent error signal and noticing that there is a unique time-dependent opportunity for cost-effective correction. Chiller and boiler replacements during renovations are common examples of situations like this.

Therefore, one would expect that a working PID controller would alert the customer to its full range of cost-effective alternatives to buying power, and that the customer would eventually exhaust the energy-efficiency opportunities and the error signal would diminish to the point that it was simply not worth the time or effort to entertain changes any longer.

Energy efficiency is the easiest customer demand-response strategy to envision, especially given the number of customers who have pursued this path using performance-contracting partnerships. One would expect that the most successful energy companies would use energy efficiency as a key differentiator in open, competitive energy markets. We believe that is a model for the future.

Let's press on to use the PID model for time-dependent price or reliability variations. As indicated in Figure 4-6, power plants used for fewer and fewer hours tend to require higher prices to compensate for the lower annual run times. Since loads tend to be weather dependent, this generally results in the highest-priced power occurring during extreme hot or cold weather. So, even though the annual hours of operation are small, the cumulative annual benefits for reducing loads during those hours are actually of similar magnitude to energy-efficiency benefits.

Once again, if a long-run avoided cost price signal was offered by an agent against which customers could evaluate their opportunities and adjust their "set points," the following PID analog would emerge. The customer (or its agent) would look at the cumulative value of each potential demand-response program and decide what portfolio of participation made the most economic sense. Given an established long-run valuation, the customer (or its agent) could make an investment decision in both automation and information systems with some confidence about the likely annual "exercised" use and economic valuation. And, as with energy efficiency, each customer would hit a law of diminishing returns in most opportunity areas either because the economic returns fell below an acceptable level or the level of participation became disruptive to the core business interests of the customer.

The economic valuation signal (the error signal perceived by the customer) would emerge from the combination of a third-party buying agent willing to take the long-run risk of reserving the customer's capability over the term of the agreement at a fixed minimum plus some pay-for-performance operating benefits. The reward for this agent would obviously be the infrequent and uncertain high valuation when regional markets went into constraint. But, without the willingness of the agent to write the long-term swap agreement against the market and offer the customer a reasonable value signal for its response, the customer would not participate and be available. The new wrinkle in this type of agreement (as compared to energy efficiency) is that there are rightful costs and risks that have to be factored into the price signals taken by the risk-aggregating counterparties since not all customer demand response is equally valuable as an alternative to supply-side options. Said very simply, the demand-response mechanism itself and its specific temporal characteristics would determine its final value, unlike energy efficiency, which simply avoids the consumption of the kWh.

The proportional element is the customer's cost-effective ability to participate in demand response. As the customer attempts to be increasingly load-shape flexible, the customer will run into the same law of diminishing returns it did when it pursued energy-efficiency options. It would therefore be only logical that the customer would only consider opportunities until the marginal costs approached the marginal benefits. Integrating the error signal would alert customers to the cumulative and interactive effects of the relatively small unclaimed benefits such as operational errors during curtailment events that cause demand-response savings to be less than achievable, or engaging in nearer-to-real-time savings improvements that require automation. It is quite likely that an integral error signal on the benefits of demand response lost due to the inadequate communication of meter information would compel customers to pay for on-line data presentment during event days. Another likely outfall from integration is the decision to automate the demand-response capability in response to price, to get the typically higher valuation of closer-to-real-time market signals.

The differential control system element would once again alert the customer to the time-sensitive opportunities to enhance its opportunity portfolio. This would include the customer's interest in an emergency generator for reliability or loss prevention (e.g., food loss in supermarkets) where the design or selection of the generator would be changed given participation in demand-response opportunities. For example, paralleling switchgear and the choice of a natural gas engine are both likely when a customer was prompted to consider adding generation for other reasons. In addition, advanced metering and energy information systems would possibly appear differentially if that added information proved to be of core business value. These opportunities are emerging rapidly in the commercial national account market as the FDA insists that food-merchandising companies keep track of and report food storage temperatures. These companies are increasingly at a cost risk in their base business if they fail to have on-site computer monitoring equipment and will be forced to throw away food if they lose power (a differential error signal). Many of these companies are going to buy generators. That decision will produce differentially attractive (in electricity markets) demand-trading capability decisions as well.

Once again, we have been discussing the PID control system from the perspective of customers who are not actually facing the time-varying market valuation for their capabilities. These customers have sold that capability to another organization willing to take the valuation volatility out of the customer's perspective in exchange for the rights to exercise that capability in the regional market. So, now let's switch to the perspective of the organization writing the swap agreement itself. Clearly, this agent would want to minimize total cumulative cash exposures and would therefore want to maximize its opportunities for value creation. The error signals it would watch are the regional power prices or reliability conditions, and the consequent commercial terms resulting from that market such as load-following agreements (where the buyer simply agrees to pay the seller some price based upon its expected load shape for the period in question), call option pricing, day-ahead and day-of ISO/RTO-operated hourly and ancillary services markets.

Since the buying agent in our example (the reserve bank or the equivalent counterparty writing the agreement) would have contractual rights to exercise the customer's demand-trading capabilities, this agent would naturally trade all available customer resources in their portfolio into and as a part of regional ISO/RTO structures to achieve full valuation. It would be natural for the swap agreement writer to choose the best buying counterparty for any resources it had, including long-term call options where the financial benefits of customer participation were clear. And, given that this agent knows exactly how each customer in the portfolio performs in response to the potential counterparty agreements, this agent would be in the position to best know how to arbitrage the aggregate value. We believe this value is significant enough that there will be near-term agents willing to underwrite the risk at acceptable premiums to acquire these resources. However, without some level of "exclusivity" in a region, no agent will risk competing with others to achieve this capability. That is why we believe there needs to be just one demand-response bank in a region – it is essentially a natural monopoly.

Now, let's look at the components in the PID control system that would be at work for this buying agent. The proportional "error" signal emanates from the combination of hourly price signals and the hours of duration. The size of the potential for arbitrage is obviously proportional to this error. And, as the price signal rises, one can imagine some kind of an elasticity curve in customer demand-response availability. However, since some customer demand-trading capabilities can only be made available with certain aggregate valuation over time (as in call options), some level of integration of this error signal must also be implemented. Said very simply, it is not good enough to know that a price will be true in any one hour, there also needs to be some level of confidence that the aggregate value over a given number of hours is available.

Therefore, the buying agent would use a proportional and integral signal to trigger its decisions to commit and actuate the constituent elements in its demand-trading portfolio. The buying agent would look at the individual hourly pattern of day-ahead, day-of, and experienced real-time prices, along with ancillary service payment offers, and determine the portfolio opportunity using these signals. Differential control in this control system would be on several time scales to alert this agent of changes in both the temporal as well as spatial elements in these prices. The reason this element of control is so important is that zonal congestion tends to break prices away from general trends markedly, as other sections of this document have illustrated. Differential control system elements warn the trading agent of potentially disruptive price discontinuities which, left unnoticed, could dramatically influence the potential value of the demand-trading portfolio valuation.

As a result of a PID control system, the buying agent would be watching the time-varying nature of these opportunities and develop an actuation plan for customer demand-response capabilities that responds to the proportional, integral, and differential error signals in this time-based wholesale electricity price-value signature.

But, the reader may ask, what happens when the regional electricity market does not have prices in it that would be of arbitrage value in any given year? And, why would anyone try to secure customer demand-response capabilities today when the outlook for price spikes seems so remote? First of all, the fact that zonal congestion is still likely has not escaped those in the know today. Intellectually, most professionals recognize the hidden risks in statistics indicating we have enough generation to prevent price spikes. They are aware that the transmission system is inadequate and bound to create price-risk mitigation opportunities. It just isn't very clear where those opportunities will exist in the electrical systems or their timing. It is very much like knowing that hurricanes occur each year, but not knowing where they will hit. Therefore, if a demand-response reserve bank can be created with adequate geographic diversity, creative agents could use this mechanism to acquire these resources in areas likely to experience some level of zonal congestion over time.

In addition, in years when regional prices fail to provide overt economic justification to actuate demand trading for market price mitigation purposes, we believe this writing agent would exercise its options to gather the performance data (i.e., building the history on which to base integral responses) necessary to know what risks and rewards it could and should entertain when market valuations returned to and exceeded long-run fundamentals. After all, it is only right that the swap agreement would naturally produce positive rewards to the agent during those times. These rewards should, over enough time, compensate the risk-writing agent for both out-of-pocket expenses as well as the time-valued risk taking in which it engages. We also believe that this information has intrinsic value, very likely much greater than that associated with the demand trades, that will be recognized and monetized over time when markets do go into constraint. One only has to look at any other form of commodity trading to see proof that those with the best underlying data do best in markets over time.

The process of analysis during low market valuation years will also permit the risk-writing agent to calmly and quietly determine the appropriate discount and non-compliance adjustment to customers if any of the demand-response options do not live up to the market requirements. For example, call options require daily minimum and total monthly hours of exercise rights to replace the offset supply-side resources. Customers may indicate they are willing to perform to get those credits, but may fail to live up to even their own expectations of performance. It is far better to know that information in times of low prices than to face the consequences when prices are high. In addition, while seemingly simple in principle, the data-transfer requirements to support the measurement, verification, and settlement of accounts can only be exercised and tested by actual use. Customers would not receive payment for actions that failed to close this value loop. This minimizes the costs to the buying agents and gives them the data they will need to be more effective market-arbitrage agents.

As the regional electricity markets continue their progress toward open, competitive models, many more illustrations of the PID model will be seen in practice. But, before readers assume that control systems are automatically the best solutions to problems, they must also be warned that all PID systems require a careful practical balance between the measurement, gain, and actuation control elements. Measurement must be trustworthy, gain must be kept in balance, and the control actuation must have a feedback system that alerts the system that the correction is always in balance with the error signal. In the screen door analogy, we certainly don't want the control to break the door off the hinges when it is released. Panic coupled with naïve politics can design systems that will do exactly that. Anyone who has experience with real-world control systems will tell you that there is much more to making this work in practice than meets the eye. However, carefully planned and applied, the PID control approach holds great promise for disciplining competitive electricity markets.

5 APPROACHES FOR DEALING WITH RISK

As noted earlier in this document, once competitive markets identify and quantify risk, pricing approaches are often developed based on the level of risk associated with serving customers. For example, higher prices are paid by those in the following situations:

- Smokers looking for healthcare and life insurance,
- Teenagers looking for auto insurance,
- Air travelers who change their flights at the last minute.

In each case, the retailer will generally produce price-risk differentiated offerings for various groups of customers in an attempt to segregate, price, and if possible, minimize risks. Customers of course respond to these offerings and will often change in some way to avoid or mitigate the impact of these prices.

In the power industry, however, the concept of customers being offered risk-differentiated retail prices based upon historical or real-time load shapes has not yet been widely adopted. One might think that any commodity with the volatility of electricity would be billed either on a real-time price or at least on a time-differentiated basis for demand and energy. Ironically, the competitive market has largely set prices that are "flat" in part due to the desire to meet customer demand and to fulfill regulatory commitments made in the past.

Where electricity customers have been offered competitive choice, the competitive energy companies have most often been forced to come up with a lower price than the incumbent while assuming the price risk in that offer. This may sound silly to industry outsiders, but as noted earlier, energy companies generally have to compete against the POLR price that does not include an insurance cost, in part because the price risk was never well defined in the historical regulated power-industry model. When these companies attempt to recover the price risk in their offerings they run out of headroom and customers won't switch. Without the price-risk premium, the customer becomes a naked (unhedged) risk.

Demand trading can help to reduce this risk. However, when forward prices in the market are soft and weather predictions seem benign, few want to take action to prepare for the opposite situation. When power prices rise again, however, remember the old adage "you can't buy fire insurance when your house is on fire."

This chapter will investigate various aspects of risk and consider mechanisms the power industry can use to deal with risk, including the concept of risk pooling. This chapter will also look at how risk will affect customers and customer-migration issues.

The Risk Pooling Insurance Model

Whenever market participants feel that losses, while infrequent, can be devastating, they naturally form risk pools. The concept can be traced back to Babylonia where traders pooled the risks of caravans crossing the desert. The most famous of contemporary origin is London's Lloyd's Coffee House (1688) where merchants, ship owners, and underwriters met to transact business. The pooling of risks is even undertaken in the speculative oil and gas exploration industry, allowing competing energy companies to mitigate drilling risks.

Pooling risk enables counterparties to level out extremes in volatility. For example, racehorses are seldom owned by one investor. The cost to acquire, train and keep a thoroughbred is too high, so investment syndicates generally buy into several horses. In the same way, most people hedge their investment risk by pooling risks under the management of professionals by investing in mutual funds. The same structure may be applicable to demand trading. Most customers need continuity of benefits and will take a reduced overall benefit rather than expose themselves to the volatility in benefits that the market itself would produce.[1]

These risk pooling insurance concepts are based upon the following:

- The loss potential is either greater than the financial capacity of any one member or it is judged as fiscally inappropriate to self-insure.
- Pool members must share a high degree of similarity, objectives, and vigilance. Each of them must possess a strong responsibility for the success of the pool. Why this requirement for a high level of commitment? Without it, the most unreliable member can benefit the most from the pool. Correctly organized, however, a pool will support all of its members.

When insurance companies establish pools, they have restrictions about whom they will underwrite. Health and life insurance are more expensive for smokers than non-smokers, and car insurance is lower for better drivers. For-profit insurance companies are always looking at the risks of underwriting customers. If someone seeking car insurance has too many accidents, they are certain to find out how these companies mitigate risks. High-risk customers are assigned to a risk pool where all insurance companies must take their share of customers or they lose their license to operate in a region. The rates offered, however, are not exactly at bargain-basement prices.

[1] In the language of the trader, someone has to write the "swap agreement" offering a fixed payment (possibly with reservation and execution compensation elements) to at least partially eliminate the volatility in the demand-response market. The swap can have a true-up at certain intervals, the same way the mortgage holder escrows taxes on your home.

The underpinnings of risk pools are extremely important in today's electricity market situations. The questions we must address are:

- **Relational control –** who can control the payment of rightful risk premiums from all parties who will benefit from demand trading,
- **Risk premium price-setting process** – should all parties pay equally (such as with "postage-stamp" analogies) or are there higher-risk situations that must pay more than others,
- **Compensation –** how do you fairly compensate those who prevent losses through demand-trading infrastructure and actions (and are they all paid on the same basis), and
- **Business operation –** who can rightfully and sustainably run this as a business?

The national dialogue seems to be in support of demand trading on a regional basis. The public benefit of pooling demand trades has broad societal value both from price-mitigation and reliability-enhancement perspectives. As noted earlier, unlike health insurance, demand-trading customers can do more than protect themselves, they can also protect others in the pool from high prices. In this regard, the car insurance model might be an appropriate analogy. In this model, everyone in the region has a requirement for coverage and those providing loss mitigation receive payment.

Defining the Required Risk Pools

Three risk pools would be needed to run an electricity market:

- Forward block,
- Hourly (day ahead and day of), and
- Ancillary services.

First, we need a forward block market pool that lets people share risk in the monthly and weekly markets. This pool already exists on the supply side and, for a while, seemed to be increasingly liquid and transparent. Counterparties seemed content with the results. Liquidity was building rapidly and capital projects were being authorized based upon the forward price signals from the regional pools. There were real questions about the ability of that market to deliver over wires that were being reorganized, but the capital markets were confident at that time that this could be eventually worked out.

The combination of September 11th, the "dot.com" bombs, and a general lackluster economy diminished load growth to the point that the forward market for electricity became soft. Long-term deals were available at such low prices that energy buyers generally found it preferable to lock in all their requirements. At the same time, fallout from the California power market, the energy trading industry, and other problems produced negative press about energy trading. As a result, company after company reduced operations or left that market completely, leading to the virtual collapse of forward market trading. Liquidity on the supply side is just not there at the moment.

With this in mind, we shouldn't be surprised at the consequences for demand trading. Since demand-side resources are primarily an economic response to supply-side options, they naturally rise and fall in "attractiveness" with the supply side. It is interesting to note that regulators and others have long recognized that energy efficiency (the demand-side equivalent of the base-loaded power plant) should be compared to long-run avoided costs.

The premises for investments in energy efficiency go beyond the savings to the customer implementing the measures – they also reflect the societal benefits of avoiding energy usage. Even in deregulated markets, public utility commissions still sometimes require regulated LSEs (typically utility distribution companies) to administer efficiency programs and invest money to implement efficiency measures. Without this outside impetus, there would be little reason for a wires company to spend money to discourage customers from using its product.

Where energy efficiency is deemed essential to mitigate risks in this forward-market pool, it has been the domain of regulatory bodies to determine the appropriate level of investment, and they have generally imposed the obligation to manage this demand-response resource on the regulated wires company. This pool has sufficient diversity in it that it can survive on a region-by-region basis because the mean time between benefits is short enough to provide the needed feedback and control signals on cost effectiveness. However, if the regulatory bodies were to try to use near-term generation costs as the feedback and control linkage, they would get trapped into the same circuitous and dangerous argument outlined earlier.

Some might argue that the regulatory process is intrinsically inefficient and more costly than finding a free-market solution for promoting energy efficiency. Even in a regulated world, wouldn't it be better for customers to self-fund the energy-efficiency measures by reducing their own bills? Paying them to do what they should do themselves is intrinsically inefficient, isn't it? Alternatively, wouldn't it also be possible to use a model where energy efficiency is the basis of a price-risk differentiator in some form of performance contract with a customer? Fortunately, we do have some successful commercial examples of this. However, the reality is that with transactions costs, lack of information, low forward prices, and other factors, most customers are reluctant to implement energy-efficiency measures and fail to maintain them. In some cases, customers do not even look at their energy bills. With relatively low incentives to reduce energy usage, the commercial value proposition often becomes too elusive for the competitive energy company to pursue. That is why many energy companies who have tried to use this value proposition in the past have had to move into building operations and maintenance to build a sufficient value proposition.

In addition, customers are sometimes suspicious of outsider claims about saving them money on their energy bills. Receiving a check for the installation of an energy-efficient lighting system is more tangible.

Risk Pools for the Spot and Ancillary Services Markets

The risk pool for the spot market is where daily and hourly adjustments could be made between those who have access to excess energy capabilities and can sell to those who need it to balance their schedule into the ISO/RTO (or whomever is going to balance the supply-demand relationships to maintain reliability). Since this is most often the domain of the ISO/RTO perhaps they would be the logical choice to set up a demand-response pool where they charge all market participants. Another possibility is for the PUC to perform this role under the demand-response agenda it already has with its regulated energy companies in many regions. (This latter approach might face some concerns about letting regulated companies work with customers in a competitive area.)

Another thought before we move on to the ancillary services market. Risk pools require the ability to assess risks, collect premiums, assess damages, and disperse benefits. They require the ability to enforce these attributes on the market. Whatever type of organization runs such a pool has to ensure that demand trades will be treated fairly and equitably in this setting.

Finally, we can consider a pool for the ancillary services[2] market that is run in each region to assure that the lights stay on. Generation companies can offer capabilities in each of these markets and so can demand trading. Now we are clearly in the realm of the ISO/RTO and it seems only reasonable that they should rely on demand trading whenever it is superior to the supply-side resources. However, asking the ISOs/RTOs to build these pools has proven elusive since they do not generally have relationships with the end-use customers, and those who do may not have adequate incentives to bring this resource to the market. Therefore, once again, establishing an external risk pooling method has value. The ISO/RTO would buy from the pool, and in doing so over time would build a forward price signal for its valuation of this resource.

The formation of ISO/RTO structures has also highlighted the importance of resource location. Even where the overall regional adequacy seems sufficient, load pockets and congestion are proving troublesome and costly. Demand-response resources can provide a way to help alleviate such congestion. In addition, they can do so in an environmentally sound manner since most of the demand-trading capability is based on reducing loads, not the use of backup generation.

The demand-trading capabilities that can be employed in the ancillary services market are numerous and include the direct load control (DLC) of water heaters, pool pumps, air conditioners, etc. Most of these resources are under programs and cost reimbursements through regulated energy companies. These companies are also questioning short-run cost effectiveness, especially in the current environment of cost cutting. Perhaps then, these companies should be paid by the ISO for this resource and the controls turned over to the ISO for use in ancillary

[2] Ancillary services may include such reliability services as voltage and frequency regulation, reserves, energy imbalance, real-power loss replacement, backup supply, and system black start.

services? We certainly aren't going to expect the free market to actuate these controls and bid this resource into the ISO, or are we? We will come back to this issue in Chapter 6.

Interestingly, all of these options are reasonable. A regulated energy company could sell off (and some have) their DLC switches to a third party who would maintain and actuate these resources and sell their capability into the local ancillary services market. The problem here is, once again, the determination of benefits and the long-run economic signals offered to customers and their agents for this capability. Ancillary services costs follow the price spikes in magnitude. Not surprisingly, ancillary services costs are low when spot prices are low, and the converse is true. If there was a third-party, market-making agent (in essence writing the swap agreement) we could also see the economics for these devices stabilized.

How Big Does a Risk Pool Have to Be?

The first reaction of most professionals to the question of pool size would be that an ISO/RTO footprint should be adequate. Other reactions might be that a zonal or regional footprint is adequate. Unfortunately, this logic may not be correct. Pools are most effective if they are large enough to use risk diversity effectively, and the infrequent and unpredictable valuation for demand trades may make anything much less than a "super-regional" or even a national aggregation counterproductive.

Ironically, when a given region of the country is considered as a pool alone, the situation is fragmented sufficiently to effectively degrade the diversity required to create the pooling effect in the first place. For example, we all know that rolling one die has an equal likelihood of any one number between 1 and 6. We also know that rolling two dice creates more of a central tendency to the possibilities. Rolling three dice improves the central tendency further. The reason the central tendency improves is the intrinsic independence of the individual dice throws.

In a similar way, we have the same problem with the statistical treatment in the likelihood of price spikes in any one year and in any one part of the country. Spikes are more likely to occur somewhere in the U.S., but not in the same place year after year (if for no other reason than the weather patterns change and that generation is being installed in anticipation of, or in response to, higher prices). For example, consider the diversity between a risk pool that included customers in the Northeast, Texas, and California. The weather is bound to be hot in one of those markets each and every year. If all customers were paying a risk premium into such a pool, there would likely be adequate continuity of benefit and expense for the pooling concept to work.

Therefore, does it make any sense that any one state or region can establish this risk pool? Yes, *if there is an underwriting pooling agent on a national level*. When regional pools are considered alone and they do not have a way to lay off risk to another diversified risk-taking agent, they are bound to be knocked around very badly by a run of bad luck. This risk could be alleviated by forming the equivalent of the Federal Deposit Insurance Corporation (FDIC) in the banking industry. The FDIC made commercial banking a reality. Its mission has been to maintain the stability of the nation's financial system and increase public confidence in the system. To achieve this goal, the FDIC promotes safe banking practices and insures bank depositors against financial ruin (up to $100,000). It does so by collecting nothing from banks it determines have the strength to eliminate the risk themselves and others in proportion to the risks they pose.

The equivalent of the FDIC in the power industry could help to enable retail electricity competition. A national risk management clearinghouse could be funded by a portion of the amounts PUCs or ISOs/RTOs collect. Funds could then be paid to balance the demand-response accounts over time when drawdowns occur. While such an activity should be a commercial enterprise, it could be necessary to "jump start" the infrastructure using governmental authority and the justification of demand response's social benefits.

The Specifics of Price Granularity Risk

When is too much information detrimental? When the difference between what you think and what proves to be true is uncomfortably different. For example, while the ISO/RTO model is moving rapidly towards a five-minute market in which all price values are considered (energy and ancillary services), only generating companies can work with these signals currently.

Traders by their very nature work in hourly increments, and even they shun individual hourly trades. Traders only resort to individual hours for same-day trades. Day ahead, they tend to trade in four-hour or larger blocks. For example, the largest reporting agencies (Bloomberg and Platts) only report day-ahead activity in 16-hour trades for the entire on-peak period, even though there are fractional day trades as well.

The reason is that there is insufficient liquidity on the supply side in the four-hour and shorter time periods for the reporting to be meaningful. When an ISO/RTO defines market rules that perturb these time periods, such as forcing day-ahead markets into individual hourly blocks, they create significant granularity risk – the chance that the individual hourly prices may not correlate very well with day-ahead pricing. Proof of this abounds. PJM has had many instances where the day-ahead prices simply do not correlate at all with day-of hourly prices when the supply-demand balance is strained. Worse yet, the "shape" of the real-time hourly prices sometimes doesn't even correlate.

Where the possibility exists that individual time increments can have significant errors, there are two directions risk management strategies can consider. One is to directly link the option to the price signal so that the only time the signal is sent is when the price is consistent with assumptions (interruptible agreements follow this logic). The other is to aggregate the time signal so that an average can be known with more certainty (i.e., less risk).

Managing Risk – the Customer Migration Process

Mature industries such as health care and insurance recognize that the economics and profitability of the business itself can depend upon the specifics of how they identify and price risks, how they manage customer portfolios, and their methods for assessing consequent losses. Anyone who has had several traffic tickets has probably learned that insurance rates only stay inexpensive when you don't exhibit risky driving behavior. Similarly, and unfortunately, people who experience life-threatening health conditions can face very high premiums.

In the power industry, the opening of retail markets promised savings, and aggregation seemed the obvious and natural way to get those savings. After all, wouldn't it be silly for most customers to go out to the wholesale market by themselves to secure an energy supplier?

The clear answer is for their loads to be added to those of others and served in aggregate. However, many large customers would not be happy with this type of aggregation. They might look at those they were being averaged in with and, if their load shape or load flexibility was better than their counterparts', figured they would be better off alone, or at least with others in a similar group. The result is that the "best customers" tend to split off from the aggregate. The remaining aggregate gets to look worse and worse over time, until the energy company can't afford to compete for what is left and, given a chance, may send them back to the POLR. This pattern has begun to emerge in electricity markets, pointing clearly to the flaws in the power industry's mechanisms for managing the customer portfolio and its migration at the moment.

In the power industry, a model of risk control might be characterized as follows:

- **Awareness:** the perception that something is different from the "average" or normal. This has certainly occurred in the electricity markets in many regions of the country, but is being thwarted at the moment by confusion about how to read price signals. There has been too great a focus on supply-side reserves as a method of price protection rather than integrating demand trades into the markets as alternatives to supply.
- **Valuation:** the ascription of good and bad attributes to what is different. This captures one of the most elusive metrics at the moment. Seemingly everyone now admits that demand response has significant value, but the performance metric is based upon "looking like and acting like a generating plant" rather than providing protection from the high price or reliability situation in the first place. Everyone now wants to compare demand response to a generator rather than to compare it against avoiding the generator in the first place. As with the arguments for the rightful place for energy efficiency, there needs to be a "long-run avoided cost" metric to determine the cost effectiveness of the demand-traded resource, not a near-real-time avoided actual price in a reliability stack.
- **Assignment:** the discrimination of retail propositions based upon these attributes. As long as competitive energy companies can survive without such considerations they will. After all, unless mandated, how many of us would carry insurance on our cars if the car itself were almost worthless? Once the first energy company succeeds with price-risk demand trading as a differentiator, competition would naturally result in others following suit. As Chapter 4

suggests, the situation at this time is a matter of the mean time between benefits for demand response exceeding the term of the retail agreement. If the benefits were annualized and "reasonably certain," energy companies would use it as a differentiator.

- **Migration:** customers attempt to shift risks by changing suppliers, while suppliers try to do the analogous activity. It is fascinating to watch insurance retailers offering to help customers compare price offers and deliberately suggesting that other retailers might be better able to handle the undesirable customers who call. Once again, the ability to use migration depends upon the elimination of the safe place for these high-risk customers to hide – the POLR. Mature markets minimize migration by standardizing the parameters around which prices are offered. This minimizes the "seams" that customers can find to migrate unfairly.

We have a game of "Old Maid" being played out at the moment. In this game, the player who has the Old Maid card in his/her hand when everyone else has matched up their cards is the loser. Customers are passed around retailers until the price spikes occur, at which time the card is passed to the POLR or the energy company serving the customer goes bankrupt.

Customer Aggregation Creates Liquidity

To understand how aggregation (pooling) creates liquidity, consider a group of customers on the ultimate price-risk shifting model – real-time pricing (RTP). Setting aside any questions about the accuracy of the price signal, let's imagine how these customers could trade demand response with others through an aggregating agent. Since these customers have complete price exposure for the loads they committed to such a pricing scheme, they might be willing to "pay someone else" to take their place on any given day.

For example, let's assume the day-ahead price forecast was for individual hourly prices to hit $1,000 per MWh ($1.00 per kWh) and the customer in question has a production situation where they really do not feel it is in their best interest to curtail loads. They could pay someone else something less than this price to curtail and thereby mitigate their price exposure much like an energy company would under the same situation. The problem is that they will have a terrible time communicating and finding a "match" for each hour of such an agreement.

However, if a group of these customers were to aggregate their needs for demand reduction, their agent might be able to arrange a match (for a fee). The agent could find the prices at which customers were willing to step in and provide demand reduction along with the resource levels needed to match volume balances. Aggregating the situation eliminates the one-for-one matching that would otherwise be so challenging.

The theoretical business model for the agent to aggregate and match up counterparties is compelling. The agent would have to cover its costs and risks of doing this, but would be offering a potentially valuable service to both the customers on RTP as well as the customers on some other price signal who would now receive some financial benefit for helping each other out. Setting aside measurement and settlement risks, and assuming a perfect market model, all opportunities like this should be able to find each other, shouldn't they? In a perfect world, they should.

As discussed earlier, price granularity can be a significant problem in power markets, and granularity defeats liquidity. Liquidity can be encouraged by standardizing the demand-response markets so that aggregation can be made to work. For example, the standards might be day-ahead four-hour blocks, day-of hourly blocks, direct load controls, and ten-minute reserves. If the market has a need, these aggregates could be expanded to two-day-ahead 8- or 12-hour blocks, and the traditional 4-, 8-, 12-, and 16-hour on-peak blocks for multiple-day commitments. The tradeoff is that the more "structured products" traded, the lower the liquidity within each, and the higher the risk. When it comes to getting customer attention, experience proves the number of products must be kept to a maximum of three or four choices for the best results.

Customer Attitudes & Choices

Where is the customer in all this? Are they looking forward to electricity choice? To a certain extent, customers are always looking for ways to save money and improve their lot, and many will switch to alternatives that will do that. Accordingly, price is a part of that decision. However, one can look at customer decisions as largely being influenced by assured savings (as opposed to theoretical or calculated savings) along with two loyalty and inertial drivers that are discussed below. Customers want to know they will save money by making choices, and the retailer who best communicates the simplicity and assuredness of those savings has the best chance in the business.

For example, when natural gas choice was offered in Georgia, several retailers talked about their prices and compared them to the prices of others. Unfortunately, there were several other bill determinants, including the peak gas consumption and monthly billing charges that confounded customers.

The first major factor regarding how customers chose suppliers was a financial one. The winning retailers were those who simply told customers they would receive a $50 check when they signed up. We could describe this as "*financial attractiveness*."

The second major factor was "*mission coincidence*" or the feeling of belonging to a like-minded group. This was achieved by aggregating groups of schools, religious organizations, senior citizens, or other types of entities that saw themselves as being of like mind and wanting to work together to be a part of something. This is akin to affiliate marketing and might be likened to an "emotional attractiveness" factor in a competitive arena. Competing industrial companies are not likely to feel this way and generally fight this type of aggregation.

The third major factor was "*embedded dependence*." This is a physical relationship to the customer that could be enabled by the installation and operation of a generator, controls, performance-based agreements, shared-savings agreements, or whatever else helps the customer save money.

These three factors interact to produce retail migration from the POLR to competitive energy companies. Where a competitor offers a price that cannot (or does not) make good business sense to any given energy company, that company has to decide to match (or beat) the lower price, lose the business, or make up for its price disadvantage in another way (by using mission-coincidence or embedded-dependence factors). The customers targeted may value these factors and be won, or they may not and would be lost.

The time constant in these propositions is extremely important. Without a mission-coincidence or embedded-dependence involvement, some fraction of customers are prone to switch annually or even more frequently if allowed. The reason for this is that many customers who switch once become comfortable with switching and will switch frequently. Those who are either afraid to switch or see little reason to switch won't switch – period.

When the embedded-dependence factor is involved, the nature of the relationship changes the time constant of the customer's decision making from the typical annual cycle to the longer period that makes sense considering the investment. The mission-coincidence factor also naturally lengthens the customer's decision cycle. Members of the affiliated group assume others will be watching out for any opportunities to save or missteps on the part of the supplier, so the intrinsic desire to switch is somewhat nullified. Mission coincidence also reduces the individual cost of customer acquisition. There may be lower-cost incentives the retailer can offer that add to the perception of value for these mission-coincident situations, such as buying the schools PCs based upon the number of customers aggregated, etc. Our observation is that 2- to 4-year horizons are created with many mission-coincidence factors.

Figure 5-1 presents a general model of how demand response could influence customers' switching behavior. This diagram assumes the POLR is not permitted to have a demand-response relationship with customers (or any other type of relationship that would substantially involve mission coincidence or embedded dependence). The "Retail Price & Program" box represents the retailer's (competitive energy company's) price offers plus any programmatic elements such as a demand response, energy audit, power quality troubleshooting, etc. Each retailer tends to select its offers based on its target market and its view about sustainable margins. A customer would compare a retailer's price and program offers to those of other retailers and the POLR price to determine the savings and other benefits. When considering this model, of course, one needs to recognize the tendency for many customers (especially residential customers) to stay with the POLR under any circumstance. Only a certain fraction of customers will even consider switching.

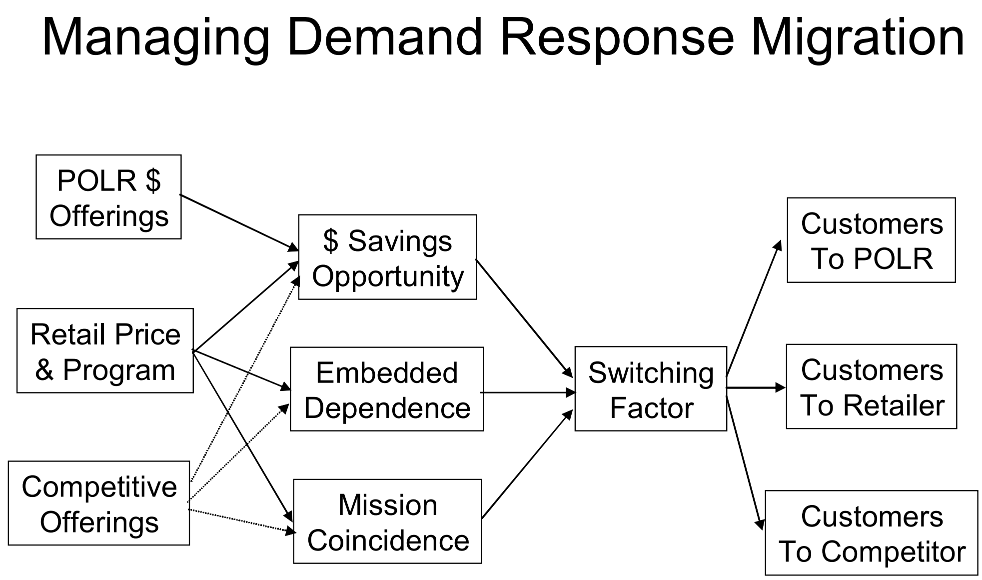

Figure 5-1
General Model for Customer Switching

In developing its business strategy, an energy company needs to determine the appropriate customers to serve in a competitive market. It must decide "who it wants to do business with" and not try to be "everything to everybody." However, that thought process seems to be stalled at the moment with the POLR in place. Once the POLR option is eliminated, customer choice will become more real and the role of demand trading will become more important as a differentiator for energy companies.

In its business strategy, an energy company should consider the fact that customer demand-trading capability is quite different from many other commodities:

- It is a relatively limited resource (it is not cheap nor easy to expand), and as a result
- Once the inexpensive resources are captured, the costs for additional resources escalate rapidly.

When these characteristics are coupled with the typical situation where the largest demand-trading resources are concentrated in a relatively few large customers, it sets up the possibility of a substantial first-mover advantage. That is, a retailer, aggregator, curtailment service provider, or other entity that initially grabs this "low-hanging fruit" of the market can preempt others from doing the same. This could provide a substantial competitive advantage, especially over time.

What Should We Learn from All This?

Markets are powerful economic forces, but thinking that they will spontaneously evolve mechanisms to counter every problem is simplistic. Similarly, thinking that simply creating a market mechanism like independent curtailment service providers will really change anything is also unrealistic. The whole market structure and the incentives along the value chain must be aligned to allow those interfacing with customers to economically acquire the counterparty relationships needed to make demand trading work. Once any one energy company can build a sustainable market model, others will certainly follow. Xerox built its business by selling copies at $.10 a page – a price people were willing to pay for the convenience it provided. Now, printers and scanners have made such great strides that the photographic film processing industry has to fear them as alternatives. We must seek and promote simplicity to create retail propositions and then let competition drive efficiency into the market.

A key question at this point is how demand-response resources should be acquired as mechanisms to promote reliability. It is interesting to note that market forces are generally not considered adequate to ensure reliability in the first place. The reason is that price communication might not be quick enough to provide adequate resources. As a result, criteria (such as some level of reserves) have generally been used to preclude the worst-case scenarios because waiting for the economic justifications can be socially irresponsible.

As noted earlier, the electricity industry is coping with some of the same criteria that the Civil Aeronautics Board does when it comes to the risks associated with airline safety. Retail air carriers must meet certain criteria before they are permitted to fly commercially. The FERC's proposed SMD contains elements that aim at this goal. However, perhaps an additional component in the interim is to consider criteria-based market solutions rather than focus on market prices.

6 MECHANISMS FOR DEVELOPING FORWARD MARKETS

Things are not going very well in U.S. power markets at the moment. Stock prices and bond ratings have dropped dramatically and the future is now viewed with uncertainty. An unintended consequence of this has been the effect on counterparty credit confidence and the commercial terms that naturally result. In general, industry participants have less trust in each other and have little confidence that power prices will return to anything near normal market valuations. Most energy companies do not have the cash to back forward positions and are concerned that their counterparties could expose them to significant financial risks if they fail in the market. In addition, for many regions the traditional systems that helped to assure grid reliability are in a state of flux as they transition toward more market-based approaches.

Linking the Disjointed Electricity Value Chain

Unlike the supply side, where the supply chain is contiguous and by necessity actively in continuous use between the buyers and sellers, the demand side is disjointed, potentially uncoordinated, and generally inactive most of the year (if active at all in any one year). The reason for this stems from a combination of factors, including the intermittence of the value proposition and the typical customer's disinterest in making energy investments for what are viewed as transient and unsure benefits. This makes it a perpetual challenge to "engage" the customer's interest.

While the long-run benefits of demand response are clear (especially considering the societal benefits), and the traditional participation venues are still correct, the intermittence of benefits is a real problem. Forecasts for cooler than normal summers and adequate regional supplies spawn disinterest and encourage energy companies to go naked. Then, when there are price spikes, the valuation becomes obvious, but few customers are engaged. Often, during the next summer, when everyone seems ready to take advantage of price spikes, there aren't any.

As noted earlier in this document, such boom-bust cycles make it difficult to maintain continuity for demand trading efforts. Customers want some level of annual benefit assurance to stay involved. There are two basic ways this can be accomplished: through regulation and free markets. The regulatory approach, which has been the traditional answer, is inconsistent with current attempts to open electricity markets to competition. It uses least-cost planning to determine the long-run avoided costs and benefits of customer participation. While a familiar alternative for many, that model would likely stall forward progress toward open, competitive electricity markets.

A free-market solution would involve a risk-taking body that understands the long-run value proposition for customer demand response and underwrites the acquisition and coordination of this resource in anticipation of future reward. This sets the stage for consideration of what might be called (for lack of a better label) *regional demand-response reserve banks*.

The Concept of Regional Demand-Response Reserve Banks

The liquidity in the U.S. dollar is maintained by the Federal Reserve Bank System. Reserves are a natural model for any resource that has extreme variation in time-based need. We have strategic petroleum reserves for exactly the same reason. And, since the reserves must be reasonably close to the point they are needed, one can envision a set of regional reserve banks, following a similar model to the way wholesale power markets are now being organized.

These reserve banks would be places where customers could deposit their existing demand-response capabilities in exchange for periodic interest payments that represent the reservation fees plus performance-based use transaction fees. These could also be places where customers (or their agents who would construct resources at their locations) could borrow money to invest in additional demand-response resources, similar to the way customers finance their homes and business investments. The reserve bank would naturally arrange for an independent audit of these resources (much like the combination of building inspectors and property appraisers) and would retain the rights to use these resources for regional electricity markets, consistent with a set of approved operational guidelines.

Once the resources were proven deposits in the reserve, their capabilities could be made available to all market counterparties: ISO/RTO organizations, retailers, generators, and even other customers (perhaps those facing undesired real-time prices). These counterparties could all purchase capabilities from these reserves and could even sign aggregate call options against these reserves with assurance that the capabilities would perform and were financially firm.

It should be noted that this concept could also apply to the supply side of the power business since it would alleviate many of the problems in forward market liquidity. The focus in this chapter, however, will continue to be on promoting liquidity of the demand-response elements of the industry.

The reserve bank would use clearing members in a manner similar to the New York Mercantile Exchange. The bank would also assure all market counterparties that the demand-response resources were not being sold beyond their proven capabilities – i.e., no gaming permitted.

Buyers from a reserve bank would naturally have quite varied interests. Some might only want the reliability-enhancement, near real-time resource benefits. Others might want capacity cost avoidance benefits, and still others might want to shift energy patterns from one part of the day to another.

Some of these concepts are complementary and some are not. There will be times when a customer's demand-trading capabilities can be bought by multiple parties during a month (or even during one day), and times when it shouldn't. There will be precedent relationships for some customers due to bilateral agreements for their resource, and times when their resource is open to see a bid from anyone in the market. There will be customers who want to exercise their options themselves and situations where customers designate others as their agents.

This concept can permit all of this commercial-transaction flexibility while assuring all counterparties that the customer's demand trades are fungible and reliable. The source of funds into the reserve bank could come from energy counterparties who want to invest (and thereby share in the reward derived by the bank) as well as pure financial interests. There may even be a "green investment portfolio" in this bank to subsidize (at least for the moment) the adoption of solar, wind, biogas, and other green energy options. Customers would be free to shop as often as they want, or stay with the incumbent energy provider, while keeping their resources on deposit for all parties.

Demand-response capabilities could be placed on deposit by customers themselves or by third-party intermediaries (similar in concept to curtailment service providers) that aggregate the demand-response resources of customers. There would also be technology-enabling partners (focused on controls, distributed generation, etc.) who would look to the reserve bank for commercial loans to finance additional demand response that can be put on deposit with the reserve system.

As in all commercial banking situations, there would be a limit to the level of investment based upon the intrinsic value of the resource. If the commercial terms were fundamentally in line with any given level of investment, and the customer was deemed a prudent investment candidate, the reserve bank would underwrite the appropriate portion of the investment. This has obvious value to customers interested in advanced technology but without either the market mechanism tolerance or the intellectual willingness to get involved in the purchase of the technology itself.

As noted in earlier chapters, the control linkages for demand response theoretically exist along a time-domain continuum similar to supply-side options. For example, investments could be made in energy efficiency as opposed to generation, and third-party ESCOs have made a business of doing this with customers. Seasonal energy management alternatives could follow the same model, once again providing a value signal based upon displaced or avoided energy costs. In both cases, the customer and the free-market agent have a clear and repeatable value proposition since the savings appear each and every year. (There may be more or less savings in any one year, but the results are fairly certain over the duration of the agreement.)

On the other hand, demand-response valuation for day-ahead and day-of markets can have years in which the value is so small that no commercial terms are viable. While the value in a year with significant price spikes can be impressive, the uncertain nature of that result is likely to discourage both the customer and certainly the free-market counterparty from placing such a bet. ISO/RTO organizations can offer capacity payments to encourage such activity, but that mechanism is certainly not widespread, and the recent FERC SMD suggests that such a model is less desirable than open-market approaches. Therefore, the concept of a commercially underwritten reserve bank is a candidate. It is conceptually possible that these reserve banks could be jump-started by the government and then turned over to commercial interests.

Figure 6-1 provides a simplified diagram of this reserve bank approach. While this presents just a conceptual overview, it offers insights into a mechanism that could circumvent the variable nature of the demand-response business and promote the various benefits of demand trading.

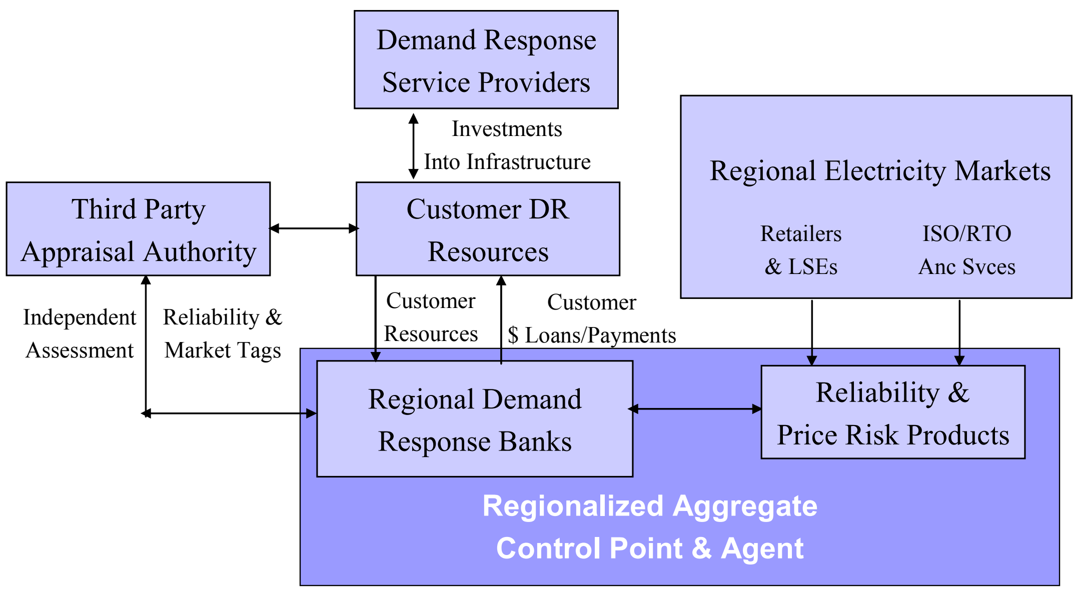

Figure 6-1
The Regional Demand-Response Reserve Bank Concept

While the functions of the reserve banks might eventually be handled by energy companies as part of their price-risk differentiation approaches, the reserve bank concept would greatly accelerate the move toward liquidity for demand-trading markets. At some time in the future, the market may be adequate enough to support a very large number of commercial banks in demand response. However, to get this market going on stable footings, some level of exclusive aggregation seems necessary. Putting all of the resources into one regional reserve bank for an area has the best chance for promoting aggregate value and liquidity. Once that was proven to work, further refinement would certainly evolve.

The establishment of these reserve banks and the associated audit valuation would permit retailers and third-party entrepreneurs to offer customers more options with commercial confidence. These would naturally include the traditional third-party shared energy savings, performance contracts, and the host of other creative energy options we see today. The mechanisms would be similar to the way local and regional banks do business. However, unlike traditional bankers, this type of reserve bank would understand the wholesale market and its nuances. As a result, the bank could design commercial interfaces to the energy retailers, generators, ISO/RTO organizations, or whomever else the reserve bank determined were rightful market counterparties.

Over time, it is expected that the reserve bank would recognize the repetitive nature of agreements and launch structured products to facilitate activities for all parties. As the market developed, the reserve bank would be in a position to look over its asset portfolio and decide what level of customer resources should be developed and where. The bank could then move toward a desired portfolio of resources (the desired number of water heaters, air conditioners, etc. vs. generators, controllable lights, or whatever) and the total resource level needed. This could help the demand side avoid falling into the same trap that seems to have occurred on the supply side (where overbuilding has destroyed the business case for many regional generators).

The reserve bank would be a natural neutral third party to market participants to develop the required resources and to act as the provider of price transparency and arbiter of fair terms. ISO/RTOs would no longer have to worry about the acquisition of this resource. They would only have to consider how the resource best interacts with their responsibilities. Energy companies could contract with the reserve bank to cover any resource requirements, and even regional generating companies might purchase seasonal call agreements with the reserve bank (as alternatives to obtaining such agreements from other generating companies).

This concept continues the ISO/RTO in its natural role as knowing best when, where, and how customer short-term notification and direct control demand-response mechanisms can fit into its reliability portfolio. And, since ISO/RTOs have no direct relationship to these customers, they will need agents and other market counterparties to acquire and financially settle with these customers.

The Devil is Always in the Details

Clearly, the general outlines presented above provide only a cursory look at the concept of a demand-response reserve bank system. The following brief discussion highlights some of the elements that such a system might include:

- **Installed Base:** The installed base would include direct load control (DLC) water heaters, pool pumps, air conditioners, dimmable lights, and whatever else can all be “reserved” at some price per year. In addition, some of these resources might be automatically actuated on acceptable financial terms while others would require the customer to bid them in, possibly in response to prices, or perhaps along with a price offer. The independent appraisal authority would assure the reserve bank of appropriate equity limits consistent with market valuation. These valuations could include economic factors such as environmental offsets, time-actuation sequence capabilities (i.e., the realistic time limits for response, total hours of response capability, etc.), and rebound effects. The independent appraisal authority could perform these economic evaluations by looking at "comparables" in the market, or even by establishing open auctions to establish market values for proposed capabilities.

- **Proven Resources:** There was a time when the U.S. dollar was backed by gold. Now, it is backed by the faith in the U.S. Government. The reserve bank would have to prove its resources, via an independent third party, for commercial terms to be viewed with confidence. The bank would naturally aggregate all regional capabilities and be sure they are each inspected at reasonable frequencies for performance. The aggregate diversity would certainly permit some level of performance risk mitigation.
- **Dispatch Agents:** Asking the customer or its selected energy management agent to know the best times to actuate its demand-trading resource is fraught with difficulties. Such organizations are not usually veterans in the regional energy markets. Market counterparties would probably be the best agents for demand-response dispatch since they are closest to the specifics of their value propositions. These agents would include energy retailers, regional generating companies who are trying to either find replacement power or cover some reserve adequacy attribute, and of course the regional ISOs/RTOs. In some cases, the regional wire company might also be in the market for either reliability or load-pocket relief. The reserve bank simply assures these counterparties that the resources are there to use.
- **Settlement Authority:** Everyone in the value chain deserves fair compensation and confidence in the fairness of all transactions. A third-party valuation provides that and a reserve bank is a natural clearinghouse. The customer would probably receive a periodic payment reflecting the reservation of its asset for use plus some performance payment recognition as it is used. If the customer had borrowed money to invest into demand-trading infrastructure, the customer would naturally pay back that investment according to the agreed upon terms. Therefore, it is quite likely that the customer could net these two against each other, and if the savings were higher than required payments in a given year, the customer might even "reinvest" these savings against an increase in debt to justify this reserve bank financing additional demand-response capabilities. This internal cash flow reinvestment may prove extremely desirable in situations where the savings would otherwise flow to the customer and get netted out against operations.
- **Market Monitoring and Oversight:** No one customer has the time or inclination to police market counterparties, and would rather not have to complain to energy counterparties about what they see as market abuse or unfair practices. A reserve bank is a natural public trust and would have an obligation to the public to retain a third-party agent to review and independently verify customer capabilities. In addition, because of the hopeful scale of these activities, the reserve bank could rightfully retain the appropriate surveillance talent – something customers would seldom be able to justify.

An Example of Risk Syndication and Reward Sharing

The following "straw man" set of economics of risk syndication and reward sharing further describes the reserve bank concept. In this model, customers (possibly sharing risks and rewards with their agents) would receive a reservation payment for proven resources plus some reward-sharing mechanism for actual use. The reward could include a minimum cost per event-day execution, a certain amount per hour of operation (if appropriate), and some recognition of the market valuation of the displaced transaction (as defined below).

A hypothetical model might be a reservation fee of $12,000 per MW of proven demand reduction per year, $500 per event day per MW for proven operation, and $150 per MWh of actual operation called for. Given that the most-likely portfolio of events in a summer when the resource was needed and used is 50 hours and 6 events, the total compensation would be $22,500 per MW. This resource would have an average implemented price of $450 per MWh (= $22,500 / 50 hrs/yr) but an incremental price at $150 per MWh. Finally, if the market cleared in any hour at a price higher than $450, the customer and its agent might be assured that the benefits would be shared 50/50 with them. This type of arrangement where cash flow assurance is offered will synchronize the interests of customers and agents.

There could also be a formula for valuation which includes market congruence (as in solar photovoltaic systems where output tends to match load-shape and price-risk exposures), ramp-down and ramp-back speeds, rebound effects, and possibly even a form of externalities to reflect any superior environmental attributes.

Asset Tagging for Transaction Matching

The reserve bank would tag a demand-response resource as to its type, its precise location in the electric system, and its impact on net load shape change (as defined in an earlier chapter). The reserve bank could then offer these resources for sale within the zones, aggregated across zones, and possibly even coordinated between zones to regional electricity market counterparties without fear that the resource was "sold twice."

Where the controls to a demand-response resource need to be physically tied to a particular market counterparty (as in certain ancillary services purchased by an ISO/RTO), the reserve bank might enter a three-way agreement with the counterparties to secure the use of the assets. This would assure all market counterparties that the resources were not under multiple operational and conflicting agreements. Anyone who has financed a large capital item is familiar with the similar document that assures lenders that they would have clear title to an asset on loan default.

Aggregate tags for water heaters are certainly appropriate rather than attempting to tag individual devices – this would be consistent with the way these assets are managed and dispatched. For example, if the controls cannot differentiate to within the zone itself, their effect can certainly be tagged to an aggregate of zones with some level of "estimated effect" noted for each zone. The buying agent (likely to be an ISO/RTO) would send the funds received for the operational use of these assets to the regional reserve bank to be cleared.

A two-part clearing mechanism will likely be necessary for demand trades, just as it has proven to be a requirement for the supply side. Final prices in an ISO/RTO or even through a regional market counterparty may not be known with precision as the event is actuated. Therefore it will probably prove expedient to estimate the benefits as the events occur and issue a final statement in the months following.

Customer-owned emergency generators could also receive allowance tags that authorize their operation only within environmental limits or when system emergencies might dictate superior social benefits by compromising air quality to keep the lights on. These asset tags would probably also indicate the fuel and any technology limits for the generator itself. For example, some generator choices take several minutes to start while others may be available within seconds. In addition, some customer facilities have dual-fueled generators and therefore these would operate at potentially quite different hourly fuel cost points. It might be deemed desirable to pay an operating cost premium to have that generator perform.

A
TRADITIONAL ECONOMIC EVALUATION METRICS

Whenever a project or an idea is considered, part of the evaluation will most often include an assessment of its economics. There are several well-accepted methods people and companies use, and they are presented below. Financial analysts use standard factors and formulas to compare "money now" to "money in the future," reflecting the time value of money.

These standard notation factors, as listed in Table A-1, were developed to compare a present value (P), with a future value (F), over time (n periods), where the interest (at i% per period) is paid or earned on a period-by-period basis. Annuities (A) represent a stream of equal payments to accumulate a future value or pay off a present value over n periods.

Table A-1
Standard Financial Notation Factors

Factor	Standard Notation
Uniform Series Capital Recovery Factor (USCRF)	A/P,%,n
Uniform Series Sinking Fund Factor (USSFF)	A/F,%,n
Uniform Series Present Worth Factor (USPWF)	P/A,%,n
Uniform Series Compound Amount Factor (USCAF)	F/A,%,n
Single Payment Present Worth Factor (SPPWF)	P/F,%,n
Single Payment Compound Amount Factor (SPCAF)	F/P,%,n

Present Worth

Present Worth converts a series of cash flows into its equivalent single cash present value at time zero. As with all the other metrics, it is a single number (not a probabilistic distribution). The magnitude of the equivalent present amount is always less than the simple sum of the cash flows due to the time value of money and any other discounting assumed to account for the perception of risk. The discounting rate is usually stated as the simple interest rate for commercial credit plus some corporately accepted risk and profit premiums. Terms frequently used to refer to present-worth calculations are: Present value (PV), Present worth (PW), Net present value (NPV), and sometimes Discounted Cash-Flow (DCF).

The present-worth (PW) method of evaluating alternatives is popular because all future expenditures or receipts are transformed into equivalent dollars now. Therefore, the simple interpretation is that the option with the highest (or the least negative) PW is the best. As simple as this sounds, and as intuitively appealing as this might be when comparing, for example, a lease vs. finance vs. outright purchase for cash, this evaluation metric can be troublesome in many ways. It is so sensitive to assumptions that it is easy for people to be misled if they do not appreciate the sensitivity of the results to those key assumptions in the analysis.

The present worth of an investment alternative at interest rate, i, with a life of n years can be expressed as P/F,%,n.

The time period over which this evaluation is performed plays a critical role in the discount factor and is one of the first warning signs for bias in the evaluation between alternatives. First of all, one should never use evaluation periods longer than the term of the agreement itself (which tends to be a year or less in retail terms) or the life of the option. This bias can be compounded when relatively low discount rates are chosen. Therefore, the biggest area for misinterpretation using the present worth factor is where both are true – the time period is long (anything over ten years) and the discount rate is low (anything less than about 12%).

This misinterpretation is common when governmental agency views about strategic investments are compared to normal corporate viewpoints. Government agencies tend to favor large, capital-intensive projects with longer simple payback periods or lower internal rates of return than the competing alternatives that might have been deemed more appropriate in corporate circles.

Use of the present worth metric is not appropriate if the customer will not persist in the relationship itself. This accurately points out one of today's problems with retail energy markets. Without an asset-based relationship to customers (such as the purchase and installation of an emergency generator as part of the retail energy supply agreement), there is virtually no way a customer will sign a ten-year agreement from an energy company. That is why it is very useful for an energy company to have a physical device (generator or control system) or at least a long-term win-win partnership (such as a performance contract) with customers to even consider this evaluation method.

Another consideration regarding the use of the PW method has to do with the lives of the assets being compared. If they are not equal or if the capacities for service are not the same, other financial metrics render a superior comparison. (For example, it would be intrinsically wrong to compare a 50-kW diesel generator with a 100-kW natural gas generator). If there are multiple factors that make each option different, the equivalent of the net present value valuation with the flexibility to compare almost any time- or size-based alternative is the equivalent uniform annual worth factor. This is something very similar in style to the long-run avoided cost calculation the energy industry has used for years.

Equivalent Uniform Annual Worth

One of the most elegant ways to compare alternatives with different lives or highly variable annual cash flows is to use the equivalent uniform annual worth (EUAW) or Uniform Series Capital Recovery Factor (USCRF) approach. It is simply a method of converting variations in yearly cash flows to an annuity equivalent. Then, in theory, you simply compare annuity values for each alternative. This is basically what car companies have done over the past few years. They realize many prospective purchasers don't care what the car costs to buy. They just want to know what it will cost per month to own. That is why many car companies have shifted their advertising to the lease payments per month, rather than even indicate what the car costs in the first place.

EUAW means that all cash inflows and disbursements are converted into an equivalent uniform annual amount, which is the same for each period. It does not require a comparison over the least common multiple of years when the alternatives have different lives. Only one life cycle of each alternative must be considered. This is also more in line with the way you would really make this decision. You might, in fact, change your mind after the option you selected wore out!

To compute the equivalent uniform annual worth (EUAW), convert the equivalent value of a given set of cash flows at a specified time into a series of uniform annual payments. If interest compounds annually, the equivalent uniform annual worth is equal to the present worth of cash flows times the uniform-series capital-recovery factor.

$$EUAW = P\ (A/P,\%,n)$$

Where P is the present value of the asset and the factor (A/P,%,n) can be computed or can be found in most financial analysis tables and converts that P from before to an annuity value for each of n periods at some interest rate.

When the EUAW is negative, it may be called the equivalent uniform annual cost. For example, suppose an asset consists of just two cash flows: an original cost (P) and a salvage value (SV), at the end of n years. The present worth of this asset is simply the first cost plus the present worth of the salvage value. Notice that P is negative in this case because it represents a cash outlay.

$$PW = -\ P + SV\ (P/F,\%,n)$$

Once you compute the present worth, finding the equivalent annual cost is easy. Simply multiply the present worth by the uniform-series capital recovery factor.

$$EUAW = [P - SV\ (P/F,\%,n)]\ (A/P,\%,n)$$

$$= P\ (A/P,\%,n) - SV\ (A/F,\%\ n)$$

While the elegance of these metrics is unarguable at describing and comparing alternatives with unequal lives, their use has generally failed to consider the probabilistic risks and business disturbances of low-likelihood, high-risk events. This formalism would be acceptable if the risk profiles among the alternatives were similar.

Rate of Return

Interest accrues on the unpaid balance of money borrowed. The rate of return (ROR) method calculates the equivalent percentage interest rate earned each year in relationship to the investment at time zero. One might also say it is the interest a bank could charge that would equivalence future returns on an investment with the cash equivalent at time zero. Rate of return is also referred to as:

- Internal rate of return (IRR),
- Break-even rate of return,
- Profitability index, or
- Return on investment (ROI)

and is customarily represented as i^*. To understand rate-of-return calculations more clearly, remember that the basis for all calculations is equivalence, or time value of money. In rate-of-return calculations, the objective is to find the interest rate at which the present sum and future sum are equivalent. In other words, find the discount rate (i^*) that produces a net present value of zero.

This metric is especially popular with financial executives who are prone to set a "minimum attractive rate of return" (MARR) or a "hurdle rate" they must see before they will approve capital budgets. The whole approach tends to marginalize people who are conservative. There can also be a computational challenge when the sign of net cash flows changes over the life of a project.

The MARR also represents the firm's profit objectives, as well as its perception of business and project risk. If the firm selects an MARR that is too high, many investments that have good returns may be rejected. On the other hand, if the rate is too low, the firm may accept projects that are marginally productive or even result in economic loss. Thus, when choosing the MARR there is a trade-off between being too selective or not being selective enough.

A traditional way to allocate a given budget among several competing projects is to evaluate each project in terms of ROR or some similar measure, and choose the set of projects that maximizes the sum of those measures, subject to the budget constraint, accompanied by an assessment of risk. As you might expect, although this procedure is widely used, it has some important deficiencies. First of all, the ROR method favors smaller projects with quicker returns. This tends to "skim" the cream and thereby capture many of the savings that would otherwise justify larger and possibly more comprehensive and strategically valuable projects. In addition, ranking on the ROR will guarantee to select the set of proposals that maximizes the total return on investment only if all the proposals are independent and there is no limitation on the money available for investments.

Payback Period

Payback period is the amount of time required for an investment to pay for itself (usually expressed in years). The most common formula for payback period is:

$$\text{Payback period} = \frac{\text{Investment}}{\text{Annual savings or profit}}$$

While some corporate managers will talk in terms of simple payback, few firms are interested in payback alone. Most managers want to earn a profit or a rate of return on their investments. This can be incorporated back into the payback period calculation by the following:

$$\text{Payback period} = \frac{\text{Investment} + \text{Required return}}{\text{Annual savings or profit}}$$

The rule of thumb was that most energy-related investments had to have a 2- to 3-year simple payback period or better to be considered valuable to most commercial and industrial customers. Now that many firms have exhausted their energy options with such returns, they are extending the viewpoint to 4+ years. It is unrealistic to think about this going out much past five years since the forecasts tend to become quite suspect.

Life Cycle Costing

Life Cycle Costing (LCC) is an economic measure that looks at the total costs of ownership over the life of the asset. It was developed to represent a long-range planner's perspective and is essentially equal to the EUAW times the life of the asset. Is that the same as Net Present Worth?

Depending upon assumptions, the answer may be yes. But, just as in any financial measure, the natural bias of the measure must be understood. By definition, Life Cycle Cost is the total cost of an item including production, modification, transportation, inventory, construction, operation, support, maintenance, disposal, salvage revenue, and any other cost of ownership taken for the aggregate of the asset's life. Theory dictates that a firm should always choose the lowest life cycle cost options. One of the key mathematical challenges is factoring in risk. Therein lies the rub –how can risk be included in the comparison of options?

What happens in all too many cases is that the analyst simply makes some arbitrary assumptions about relative risks between projects such as the following: Project A is riskier than Project B and has been discounted by an extra x percentage points. While this may sound precise, it fails to inform the management team about the sensitivity to these judgments.

Establishing the MARR

Top management knows that the "wish list" of capital projects in their organization probably exceeds their available funding capacity. They also want to be fair across organizational lines. After all, production always asks for more equipment, accounting may need new computers, and the facilities department may need a new boiler. As a result, it is natural for management to establish a fair rate of return or "hurdle rate" that all projects must exceed before they will consider them for adoption. Therefore, meeting or exceeding the hurdle rate is often a prerequisite for capital funding but not necessarily a guarantee of project funding.

A corporation accumulates funds from two major sources – debt financing and equity financing. Debt financing represents capital borrowed from others that will be paid back at a stated interest rate by a specified date (possibly over a period of years). The lender takes no direct risk on the repayment of the funds and interest, nor does the lender share in the profits of the borrowing firm. Debt financing includes borrowing via bonds, mortgages, and loans and may be classified as long-term or short-term liabilities.

Equity financing represents capital owned by the corporation and used to generate revenue. Equity capital comes from owner's funds and retained earnings. Owner's funds are classified as funds obtained from: (1) the sale of common and preferred stock for a public corporation, or (2) company owners for a private (non-stock-issuing) company. Retained earnings come from the after-tax profits of the firm. This is money left over after the company pays dividends. The company retains these funds for future investment and expansion.

The MARR represents the overall cost of capital of the firm plus an adjustment for risk. The cost of capital is simply a weighted average of the costs of debt and equity and is sometimes called the weighted average cost of capital (WACC). The MARR is then set relative to this cost. Sometimes MARR will equal WACC, but usually MARR will exceed WACC depending upon the risk of the project and availability of funds.

A risk-free return is often considered to be the rate offered by U.S. Treasury bills. The estimated return on a proposed project may be any value, but only those above the MARR will actually be considered for funding. The MARR typically varies from one project to another and through time because of factors such as project performance, tax structures, and capital markets.

Conclusions

All of these financial metrics can be representative ways of assessing investment opportunities when the options are somewhat similar in their risk profiles. However, when one of the options is to shelter the organization from risk itself, and that is being compared to other options of taking on more risk, these metrics tend to mislead an organization.

B
OPTIMIZING TIME-DEPENDENT CUSTOMER DEMAND-RESPONSE BEHAVIOR

The demand-response capabilities provided by any given group of customers is a complex, non-linear resource. It is not an engineering variable or even analogous to the supply side of the business where costs and returns on equity can be clinically defined. There is no production modeling equivalent for the demand side.

One can contract for certain attributes that are relatively predictable using reliability-based curtailable and interruptible agreements, but even these can prove to be somewhat uncertain when exercised fully. Customer participation in price-response trading approaches such as real-time pricing and demand bidding are generally even more uncertain. The demand trades for any one day can depend on the weather, price offers in each hour, day of the week, time of day, prior participation in trading events, and the customer's monthly production schedule (just to mention a few factors). As a result, it is difficult to predict the load reduction available from these programs under a particular set of circumstances on any given day. In addition, when manufacturing customers are "exercised" on any one day, they may need the next day to recover and build back process inventories to be able to fully respond to price signals. Moreover, exercising the rights of curtailment programs day after day can have significant customer service implications such as "participation erosion." This uncertainty complicates the lives of those who plan and operate demand-response programs.

A sustainable demand response strategy is important because for most participants the demand-response opportunities will be weather related (heat storms) and these are rarely single-day events. The impact of the heat builds over time and the consequence builds as well. Buildings heat up, power plants get stressed, regional cooling water supplies can run low or overheat, and transmission circuits get overloaded. The result is that the highest prices tend to be during the third and fourth days of heat storms. Therefore, without careful planning, too much of the demand-response resource can be triggered too early in the multi-day event series, resulting in less load reduction available at the highest-value (and possibly most prone to system failure) times in the heat storm.

Fortunately, these heat storms do not require a decision for all days of the storm beforehand or on the first day. One can develop a sequenced decision model that attempts to optimize a strategy going into the storm and then adjusts as the experience in the heat storm accumulates. For example, if the weather forecast was incorrect and the storm's impact is worse than anticipated (most likely because of a coincidental large power plant outage) the model could be adjusted to re-optimize a "mid-course correction." In addition, as time moves on and demand-trading agreements are exercised along the way, the model can decrement the number of remaining hours and events available. Therefore, each heat storm has to consider the history of prior heat storms already experienced and the actions taken.

The reasons for this decision cascade are in part due to the multitude of agreement types at work and the limitations they impose. Interruptible agreements frequently limit the total hours and number of times a customer must respond. Limits are typically imposed on the number of hours a day, the number of days a season, etc. Even residential air conditioner direct load control programs have limits on the number of hours and days in a row they can be activated before customers will call and ask that they be disconnected.

This appendix outlines how to approach these issues and illustrates some typical customer behaviors. It then indicates how to organize and build a model representing your situation and the types and quality of results this analytical framework will likely produce. This should help program planners and demand-trading agents better understand the dynamics of customer demand response from demand bidding, curtailment, and interruptible rate programs. In addition, the illustrative example here should help energy companies identify the market research, load research, and program planning needed to support such a model for their own use over time.

In addition to building a computer representation of a real-world decision system, optimization requires the programmer to define an objective function that the model is trying to minimize or maximize. That is, given the decision rules that constrain choices, how should tradeoffs be made between options, and when have you reached the best combination? It is unnecessarily expensive and tedious to build models that can answer any question. Therefore, be sure you know what you (or your management) are trying to do with the potential demand-trading resource inventory in the first place:

1. What is the goal of your management's team for the demand-trading program?
    a. Market price protection (or mitigation)?
    b. Competitive differentiation?
    c. Market price signal for installed generation?
    d. Reliability enhancement and system stress reduction?
    e. Delaying tactic or a substitute for a wires upgrade?
2. How are you going to measure economic impacts?
    a. Theoretical profit-making or loss-reduction opportunities?
    b. Block trades against actual schedule and markets?
    c. Long-run expectations or heat storm by heat storm results?
3. Do you have buy-in that your economic valuation method is acceptable?

The last question is not meant to be argumentative. Any economic optimization, at best, answers the question as posed. If an energy company's view of saving or making money changes with its exposure to forward and spot markets (as earlier chapters indicated often happens), the strategy will fall out of that. Otherwise, the economic optimization will assume the traditional arbitrageur viewpoint that all savings are worth their face-value consideration. The discussion of modeling concepts presented for the remainder of this appendix assumes this perspective. However, it is extremely important for readers to understand that this may not represent their organizations' views of demand-response opportunities.

This appendix illustrates a linear input-output approach. The easiest model of this type to build is one where the demand-trading resource is already in hand and the only question is how to dispatch it. We would suggest the reader start there even if the assumptions about customer resources are only best guesses. One could argue that this is possibly more instructive anyway because you could then "tinker" with the rules engine to see how you should design the agreements even before you go out trying to sign up customers.

Modeling Approach and Tools

Modeling is exciting to most programmers. However, when the management that wants the answers is not specific about what is needed, resources are generally wasted. It's like hearing "I don't know what I want, but I will know it when I see it." Those are pretty scary words to any programming group. They want a detailed work scope and deliverables, not something vague and subjective. This type of project will often precipitate a model design that brings out the detail-oriented people in the organization who will delight management with their knowledge but terrify all with the daunting data required to answer the question. The best approach is to start simple and add detail as it seems to add value.

This is no different from any other engineering question. One needs to make simplifying assumptions, and knowing when those are appropriate and not is part of the engineering judgment needed here. Knowledge should simplify models – not make them more complex than they need to be. And, given that your organization is going to invest in this effort with its own financial resources, we are assuming this is a serious investment you are making – not an academic, research, or demonstration exercise. Therefore, we are assuming you are trying to build a model at the lowest cost consistent with achieving the objectives.

Therefore, we suggest that you initially build the model in Microsoft Excel with deterministic variables (single-value input assumptions) and using the "Goal Seek" and "Solver" under Tools that come as add-ins with Excel. We suggest that the model be first built with explicit assumptions about the length of a heat storm and the intensity on each day of that heat storm as inputs, and the portfolio of customer demand-trading options evaluated parametrically until the modeler feels the results make good common sense. Doing it this way will also permit the required checks that the math in the customer agreement rules engine are working correctly.

At that point, it would be highly instructive to introduce the variability of some of these inputs by either using a purchased add-in such as Crystal Ball (www.decisioneering.com), or building the model from scratch in a software tool like Analytica (www.lumina.com). While EPRI makes no explicit or implied endorsement of these software tools, both of these programs are reasonably priced for the job at hand. Of course, you could simply develop random number generators for the preferred variables in the model yourself and single step a simulation using recalculation. One of the nice features of doing that is that the model becomes very useful when trying to explain strategies and alternatives to others in your organization.

Some readers may wonder why we suggest building the model from scratch. The reason is that, once the explicit relationships are defined, modeling software like the tools used in this project can enable a much more concise visualization, documentation, and analytical summary of the underlying interdependencies in a model. The graphical portrayals of the modeling concepts shown later in this appendix were developed in Analytica.

Required Modules and their Interactions

As in any analysis process, it is probably best to start with a simplistic model with just a few inputs and add detail as it becomes available and is deemed valuable. Otherwise, the model will be elegant but unsupportable in real life. The model offered here is broken into obvious modules.

Weather Module

The weather module would characterize the way heat storms occur in the area. Historical data is extremely helpful here and can be generalized by the expected number of heat storms in a season (an integer variable), their duration (in hours), and intensity (degree days). This module should have two outputs: a predicted weather pattern and an estimate of the "actual vs. prediction" variation that can be expected. That is, to what degree of accuracy can heat storms be predicted? We would suggest the modeler spend some time with an atmospheric scientist familiar with the region in question so that the "patterns" for both the weather as well as the variations in the weather are modeled correctly. For example, the weather patterns in the Southeast are quite predictable most summers. Take the weather in Dallas on any one day and that will be the weather in Atlanta the next day. On the other hand, the weather across the Northeast can be quite unpredictable and variable over a relatively small scale of geography.

The expected duration and intensity of a heat storm is usually known several days in advance, and where the experienced variations from that forecast are relatively small, it is likely that the final model will produce intuitively appealing results. However, where the opposite is true, the model will look more like a financial options model with a portfolio price-risk mitigation answer.

While it is obvious that the heat storm can start and end on any day of the week, the impact on the system should reflect that. For example, it is extremely important that heat storms starting on a Saturday in the Northeast are treated differently than those in the Southeast. Commercial building owners in the Northeast seldom run their air conditioners when the buildings are unoccupied, and therefore have to "recover" the building on Monday mornings. Most air conditioned buildings in the Southeast run the air conditioning all weekend long, or they wouldn't be able to recover the building. That is why some of the worst price spikes in the Northeast occur on the first working day following a weekend.

Fortunately, heat storms do not surprise energy traders. Almost all energy professionals watch the weather and subscribe to sophisticated weather-monitoring services. The result is that heat storms are often forecasted days in advance, and the length and intensity are usually well known. An exception to this is when thunderstorms can spring up seemingly out of nowhere and dash the heat to the rocks, surprising everyone.

Market Price Module

It is also an imperative to study the price-spike coincidence to regional weather parameters. Price spikes in the Southeast (SERC) are usually a reflection of extreme heat in the ECAR region. After all, it is always pretty warm in the South during the summer months. In fact, the "peak coincidence" of price spikes with native peaks is only about 30% in the Southeast. That means that, more often than not, prices will spike when the Southeast is NOT at a seasonal peak condition. One might model this by modeling prices separately from loads in the Southeast. That would not be a good model for the Northeast where prices spike on the basis of regional peak load conditions (or an unplanned equipment emergency).

You might dismiss the price spikes caused by historical equipment emergency situations since they are generally cleared up within a day or so. Focus on the historical prices during heat storms, since these are the situations that require strategy. Gather up day-ahead, 16-hour prices as well as day-of hourly prices if they are available. Use actual hourly data and do not use individual zonal prices unless you believe they reflect your situation. Use the regional hub prices instead.

Also, be aware of the impact of price caps on the shape of the price signal for any given day. Where prices are uncapped, or do not hit the cap during any one day of a heat storm, prices tend to rise very quickly in the first hour or two of the afternoon. They also fall off as the evening approaches, and the highest hourly values tend to be during four-to-six hours of the afternoon. Where prices are capped, the cap can be reached during almost any number of hours a day on the third or fourth day of a heat storm. It is entirely possible that all sixteen hours of the on-peak period could be pegged at the cap if there is inadequate capacity available. Obviously, this is a misleading optimization situation and highly disruptive to demand trades for reasons mentioned earlier. Also, be careful about modeling heat storms that disappear dramatically, as they tend to do when a thunderstorm quells the situation instantly. Look through historical records and model this as it occurs. Weather data may not be available hourly, but even three-hour data records will show the cliff-like changes.

Agreement Module (and Associated Rules Engine)

Customers sign agreements to participate in demand-trading programs, and the parameters in those agreements that affect optimization include the obvious advance-notification requirements (how many hours must you notify the customer ahead of the event), as well as the limits in hours per event, days per season, etc. Make your life easier by placing these parameters in a prominent "input" area so that "what if" scenarios can be assessed. You should also create "software switches" that can activate and de-active groups of customers on each day of a multiple-day heat storm. For example, we would suggest simple "active" and "inactive" switches you can turn on and off as the heat storm progresses. Do not bury the switches in the logic itself or you will have trouble explaining how the model is working.

Where data is available, you should also include the reduction that occurs when agreements are exercised on multiple days in a row. For example, most energy companies have learned that curtailment agreements almost never produce the "nameplate" MW reductions when called. The rule of thumb we hear all the time is that 75-80% of the contracted capacity is seen on the first day. Subsequent days of exercise tend to see even less, and three- or four-day heat storms can see the curtailed amount drop off by 50% or more. As the reader may have already guessed, a simple input-driven "decay" in resource with explicit assumptions is a good place to start. Then, the modeler can add personality to these characteristics as the data to support that is available.

Elasticity Module

Some things seem intuitively obvious, while others can confound even the most analytical of minds. Moreover, anyone trying to predict what customers will do in response to price signals is about to meet the "emotional generator." Customers do not offer demand response linearly in response to price, repeatedly in response to price, or anything else. They can ignore high price signals one day and respond to low ones a week later. They make up their minds and that's that. Energy companies have appealed to customers "for the common good" and achieved hundreds of MWs of load reduction several times a year and customers didn't even expect payment for it. It was viewed as the right thing to do. Of course, if you call on customers several days in a row, this philanthropy comes to an abrupt halt.

And, do not assume what customers say they will do in response to price is the same as what they will actually do in response to price. In fact, don't trust anything the customer says they plan to do – find out what they do! We were surprised by the number of days customers would reduce operations in a row at prices as low as $.10 per kWh credited to their bill when these same customers refused to sign up for curtailment programs. Customers wanted the freedom to decide if and when they would respond to price. When they had this freedom, customers did respond to price far more often and at lower economic thresholds.

We were also surprised how many customers voluntarily respond at prices less than $0.05 per kWh – numbers lower than their average electric rate. Why would they do that? Were these free riders? No, they were honestly trying to save a buck. Obviously, you get more demand response as the price offer goes up, but the amount is non-linear as was documented in the ***Demand Trading Toolkit*** (EPRI 1006017) published by EPRI in 2001.

For any given customer segment (i.e., industrial, commercial, agricultural, municipal, and residential), a model of how customer demand reduction depends upon price is a function of at least these three interacting attributes:

- The "technical" elasticity potential of the customer segment (how the segment responds to price with no restrictions, or disablers),
- How the history of demand-response trades leading up to the time in question affects that potential (the memory the customer has of previous demand trades and their view about that experience), and
- The time of day, day of week, and week of the month "discount" that should be considered in demand-trading behavior.

These characteristics have been long known among energy companies. Demand-trading programs are designed reflecting these attributes. For example, several programs have "alternative day" participation where the customer will not be called two days in a row. Half of the customers are signed up for odd days and the other half for even days. Accordingly, since heat storms tend to be 3-4 days long, the "pain" of participation is shared equitably.

Interestingly, large industrial customers are thinking about this as well from their own perspective. There are now software systems and consulting services that will help a large industrial firm rethink its entire production approach to create demand-trading flexibility. Therefore, one could suggest that the largest industrial customers can be "optimized" themselves to be better demand-trading partners. In fact, one large energy company has integrated this concept into its retail proposition. It is the right strategy for certain large energy-using customers and the results can be impressive.

Putting the Model Together

Once component modules are proven to behave according to experience or common sense, they can be linked to simulate interactions. Figure B-1 summarizes the price module's interactions with the weather module.

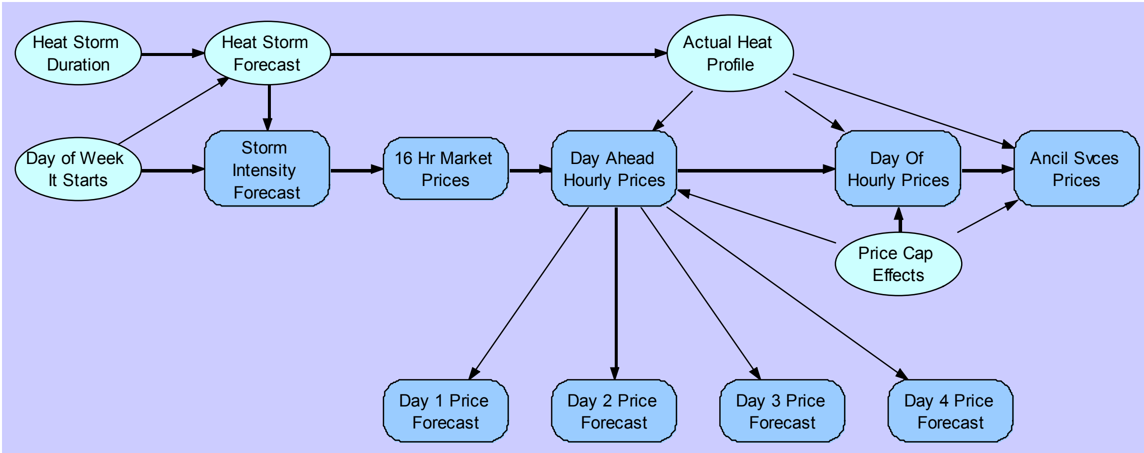

Figure B-1
Overall Optimization Model Flow

Notice that the model illustrated here specifically outputs prices for each day of the heat storm (a storm lasting up to four days) and has no feedback from actual customer demand reductions to price. That is a reasonable starting point because most real-world situations at the moment put the customer in a price-taking mode and have too little demand trading to influence clearing prices. While this might anger some who insist that customers will affect cleared prices (and they do), the reason we model it this way is that the trader usually offers the price to the customer based upon their net financial impact, often with an assured minimum of offer or actual cleared incremental price.

The next step is to build a time-stepped dispatch model that overtly selects and de-selects customer participation decisions so that all can see when and how customers are either offered price signals or programs are exercised in response to price signals. We liken this to a series of decisions a real person would make based upon the facts at hand and estimates of what is likely to happen in subsequent days. Obviously, this all starts with a Day-1 dispatch strategy.

The results from that day would feed forward, along with revisions in the weather modules shown earlier, into the specifics for Day 2, etc. Sequencing the decisions for each day of the storm follows the way energy companies would actually do this in practice. We have simplified the situation in Figure B-2 to show information flows.

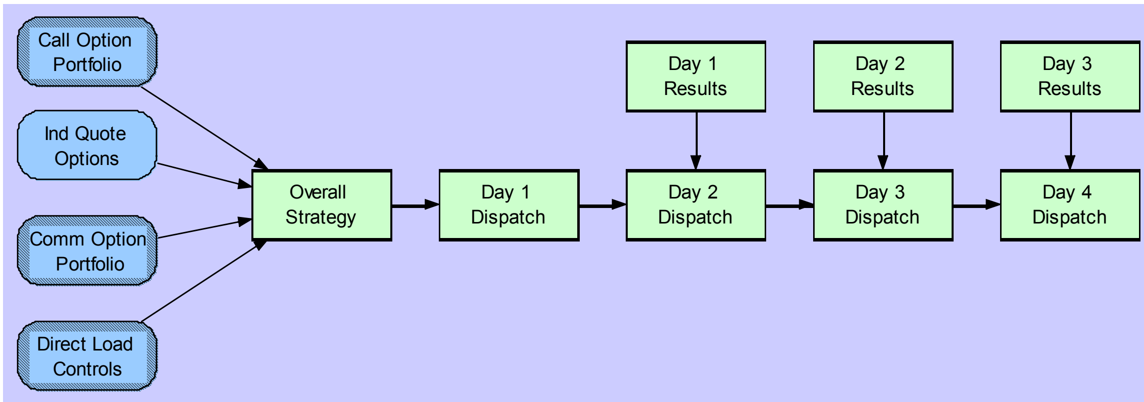

Figure B-2
Time-Sequenced Customer Dispatch Model

Once these interactions produce the correct numerical results for the expected deterministic inputs, the model should be "limit tested" to be sure that no rules are broken or unrealistic results are experienced. We also suggest demonstrating the model to those who have experience with customers to see if they spot any underlying anomalies or biases in the formulations. At that point, the model is ready for randomization and strategy development.

Goal Seeking and Optimization

At this point, it will be essential to codify the optimization rules themselves. As we indicated earlier, these are the goals and tradeoffs in the search for the best strategy. The best way to perform this step is to begin with the end in mind: what does your management want to achieve in real life? What are they afraid of (the goal here is to minimize that), and what is it they most wish would happen? Don't try to trade these off against each other just yet – just define them.

Then, when you can list the attributes for success and failure, develop a composite index that would "grade" the strategy. Corporations are doing this in the most complex of situations as they look at multiple and potentially conflicting goals and objectives. One of the most popular methods in use today is the "balanced scorecard." While the ranking may sound too subjective to be of value to some, remember that most companies need to build consensus by at least attempting to rank risks and their consequences.

Table B-1
Example of Portfolio of Customer Call Options

Option A (June through Sept)		**Maximum of 12 Calls**	
Premium ($/kW - Summer)	**$28.00**	**$24.50**	**$15.00**
Call Option Strike Price	**$0.15**	**$0.50**	**$1.00**
Hours of Reduction per Call	**8**	**8**	**8**
Energy Credit per Call per MW	**$1,200**	**$4,000**	**$8,000**
-- Total Likely $ Benefits --			
Very hot summer -- $/MW	$42,400	$48,500	$39,000
Normal summer -- $/MW	$37,600	$36,500	$23,000
Cool summer -- $/MW	$30,400	$24,500	$15,000
Option B (July & Aug Only)		**Maximum of 8 Calls**	
Premium ($/kW - Summer)	**$21.00**	**$18.50**	**$12.00**
Call Option Strike Price	**$0.15**	**$0.50**	**$1.00**
Hours of Reduction per Call	**4**	**4**	**4**
Energy Credit per Call per MW	**$600**	**$2,000**	**$4,000**
-- Total Likely $ Benefits --			
Very hot summer -- $/MW	$25,800	$28,500	$24,000
Normal summer -- $/MW	$24,600	$24,500	$16,000
Cool summer -- $/MW	$22,200	$18,500	$12,000

The following discussion illustrates how an objective function might be formulated for a competitive energy company concerned about minimizing losses when forced to offer a fixed price in a competitive market, and who already had a fully hedged portfolio of forward agreements. That is, the only reason they traded demand was for the savings it offered. The energy company had a portfolio of all customer segments and a 50/50 split of prepaid call options and price-responsive agreements. The call agreements had two options (two and four months), three strike prices ($0.15, $0.50, and $1.00 per kWh), and the "expected" total customer benefits shown in Table B-1.

One can now imagine that, depending upon the number of heat storms, their duration, and the way these options are exercised, there will be superior and inferior actuation strategies. However, while an argument might be made that the model should select the best portfolio and strategy, doing so will bait the natural questions about alternatives and their relative value, along with a forced dialogue about why the differences should be believed.

Therefore, we might suggest that the model answer explicit sets of situations with comparative costs and benefits. In this way, the relative sensitivities will be kept in focus. That is, whenever a model of complex systems is evaluated, there are bound to be some inferior strategies that the analyst can point to as such. In addition, there will be a few good strategies of relatively comparable economic value. It is imperative that organizations avoid the inferior ones, and then select the most marketable of the good ones for implementation. Said another way, theoretically perfect trading programs are probably not "salable" to most customers.

One final caution is in order. Always remember that models like this create strategies that will prove to be right in some real-world situations and wrong in others. Over a number of like events, strategy models like this tend to yield better results than just giving up and hoping. However, life has its way of surprising us all. Learn what you can from modeling by having the experienced people in your organization work on it and interacting to synthesize the best judgment. Be mindful that the best models are always developed by the best insights into life, not the most elegant and most complex of algorithms.

C
ABBREVIATIONS AND ACRONYMS

The energy industry, like most others, has the tendency to use a great many abbreviations and acronyms to facilitate communication. While an individual may understand many or most of them, there are occasions where some are not known. Here are a number of those used in this document and in other energy industry publications.

API	American Petroleum Institute
ATC	available transfer capability
bcf	billion cubic feet
BPA	Bonneville Power Administration
Btu	British thermal unit; heat required to raise 1 pound of water 1 degree F
CEO	chief executive officer
CFO	chief financial officer
CIO	chief information (IT) officer
C&I	commercial and industrial
CPP	coincident peak pricing
CTC	competitive transition charge used to recover stranded costs
DG	distributed generation
dkt	dekatherm = mmbtu, approximately = mcf
DOE	Department of Energy
DSM	demand-side management

ECAR	East Central Area Reliability coordination agreement
EEI	Edison Electric Institute
ELCON	Electricity Consumers Resource Council
EMC	Electric Membership Cooperative
EPA	Environmental Protection Agency
EPRI	Electric Power Research Institute
EPSA	Electric Power Supply Association
ERCOT	Electric Reliability Council of Texas (not all of Texas)
FERC	Federal Energy Regulatory Commission
FRCC	Florida Reliability Coordinating Council
G&T	generation and transmission
GRI	Gas Research Institute
GWh	Gigawatt hours = 1,000 MWh
ILEC	incumbent local exchange carrier
IOU	investor owned utility
IPP	independent power producer
IRP	integrated resource planning
ISO	independent system operator
kV	kilovolt
kWh	kilowatt-hour
LADWP	Los Angeles Department of Water & Power
LDC	local (gas) distributing company
LHV	lower heat value of a fuel

LMP	locational marginal pricing
LSE	load serving entity
MAIN	Mid-America Interconnected Network
MAPP	Mid-Continent Area Power Pool
MAPSA	Mid-Atlantic Power Supply Association
mcf	thousand cubic feet
mmbtu	million btu; generally equal to mcf
MW	megawatt; 1 MW = 1 million watts
MWh	megawatt-hour
NAESB	North American Energy Standards Board
NARUC	National Association of Regulatory Utility Commissioners
NERC	North American Electricity Reliability Council
NLSC	net load shape change
NOPR	notice of proposed rulemaking
NPCC	Northeast Power Coordinating Council
NRECA	National Rural Electric Cooperative Association
NYMEX	New York Mercatile Exchange
OASIS	open access same-time information system
OMB	Office of Management & Budget; a federal bureau
PJM	Pennsylvania-New Jersey-Maryland ISO and reliability region
PLMA	Peak Load Management Alliance
POLR	provider of last resort
PSC	Public Service Commission, a state agency

PUC	Public Utilities Commission, a state agency
PUHCA	Public Utilities Holding Company Act
PURPA	Public Utilities Regulatory Policy Act
PX	Power Exchange (former California trading center)
QF	qualifying facility (generation) under PURPA
RTO	regional transmission organization
SEC	Securities & Exchange Commission
SERC	Southeastern Electric Reliability Council
SMD	Standard Market Design, a proposed FERC rule
SPP	Southwest Power Pool
T&D	transmission and distribution
tcf	trillion cubic feet (gas)
therm	tenth of a mmbtu; 100,000 Btu (gas HHV)
TLR	transmission line loading relief
TOU	time of use (an electric rate)
WECC	Western Electric Coordinating Council (formerly WSCC)
WSCC	Western Systems Coordinating Council.

Target:
Market-Driven Load Management

About EPRI

EPRI creates science and technology solutions for the global energy and energy services industry. U.S. electric utilities established the Electric Power Research Institute in 1973 as a nonprofit research consortium for the benefit of utility members, their customers, and society. Now known simply as EPRI, the company provides a wide range of innovative products and services to more than 1000 energy-related organizations in 40 countries. EPRI's multidisciplinary team of scientists and engineers draws on a worldwide network of technical and business expertise to help solve today's toughest energy and environmental problems.

EPRI. Electrify the World

Printed on recycled paper in the United States of America

1001635

EPRI • 3412 Hillview Avenue, Palo Alto, California 94304 • PO Box 10412, Palo Alto, California 94303 • USA
800.313.3774 • 650.855.2121 • askepri@epri.com • www.epri.com