Demand Response and Electricity Market Efficiency

Customer response is a neglected way of solving electricity industry problems. Historically, providers have focused on supply, assuming that consumers are unwilling or unable to modify their consumption. Contrary to these expectations, customers respond to higher prices that they expect to continue by purchasing more efficient appliances and taking other efficiency measures, a review of published studies indicates.

Kathleen Spees and Lester B. Lave

I. Introduction: The Importance of Price-Responsive Demand

Edison faced the engineering challenge of satisfying the demand of his wealthy customers; he didn’t try to change their use patterns. Subsequent producers maintained the focus on reliable supply and considered shaping demand to be outside their purview. Suppliers have viewed the hourly, daily, and seasonal fluctuations of demand as facts of life. These fluctuations required additional generating capacity, particularly peaking plants that were needed only a few hours per year. Under regulation, the cost of peakers was spread over all kilowatt-hours generated, adding little to the average cost of producing power, and thus its price.

Market restructuring turned an irritation into a major problem. Independent system operators (ISOs) and regional transmission operators (RTOs) determine the price in an auction market with all successful generators paid the locational market clearing price (capped at
$1,000/MWh in all but one market\(^2\). All generators receive this price, from a baseload nuclear plant generating power at $20/MWh to an expensive light oil generator at $240/MWh (which operates only a few hours per year).\(^3\) Baseload plants can earn high profits during the high demand periods in a competitive market, but the highest-cost peaking unit only receives its marginal costs and cannot cover its fixed costs, since even the highest price is the marginal cost of a peaker.

One way to recover peaker costs proposed by economists such as Bill Hogan is to remove all price caps and allow high prices.\(^4\) High prices in California during 2000 attracted new generation investments that came online after about a year. While the high prices did attract new capacity, the investor-owned utilities, California Water Department, and ultimately the ratepayers paid manipulated high prices that persisted for more than one year.

A serious problem with the deregulated market structure is that the systems operator creates an auction market where demand is completely unresponsive to price and all successful generators are paid market price; this market design offers an all but irresistible temptation for generators to manipulate the market, sending prices soaring.\(^5\) Recognizing this problem has led to intense “market monitoring” to ensure that generators behave like purely competitive firms, bidding their generation into the market at out-of-pocket cost. The cost and effectiveness of market monitoring can be problematic and by 2002 the California ISO (CAISO) had convinced the Federal Energy Regulatory Commission (FERC) to allow a cap of $250/MWh, increased to $400/MWh in 2006.\(^6\)

Despite the occasional high profits earned by some generators, some areas have experienced inadequate investment in new capacity. The Pennsylvania–New Jersey–Maryland (PJM) RTO, ISO New England (ISO-NE), and New York ISO (NYISO) have created capacity markets to pay for fixed costs.\(^7\) Capacity markets are also problematic but are not the issue here.

The systems operators pretend that customers cannot or will not alter their electricity use, no matter how high the price. Thus customers face a fixed retail price, e.g., $0.10/kWh,\(^8\) even when the wholesale price hits its maximum of $1/kWh. A customer has no reason not to use an electric dryer at 5 p.m. on the hottest day in August because she always pays the same $0.10/kWh. If the customer faced a price of $1/kWh, she would demand much less electricity. We conjecture that, once consumers have the technology to respond in real time, the delivered price of electricity would never exceed about $0.30/kWh.

In agent-based simulation and experimental auctions, generators in a competitive auction market facing a fixed, vertical demand learned to drive the price to the cap without conspiring to raise price.\(^9\),\(^10\) When large customers bid into the market, rather than being represented by the ISO as a fixed demand, the customers were able to erode or destroy the market power of suppliers.

A regulated utility has much to gain from having customers respond to RTPs, as Schweppe noted.\(^11\),\(^12\),\(^13\) Schweppe’s vision of a dynamic demand-side electric marketplace has failed to materialize, even though an active customer role is even more important in the restructured market; customers need the ability to refuse purchases when the RTP is higher than they are willing to pay. Industry restructuring has breathed new life into demand response and generated a wide range of demonstration projects and pilot programs.\(^14\) Many market operators in the United States have developed initiatives to invite demand into the...
marketplace, but enrollments have been small and sluggish. Market operators publish lists of private parties who provide demand response services, but only a few end users currently employ these services.\textsuperscript{15,16} We explore the obstacles that public regulators and private ventures must overcome before they can transform the industry.

II. Conservation Initiatives and Effectiveness

Electricity conservation policies since 1975 have been expensive but cost-effective. A recent Resources for the Future (RFF) retrospective estimated expenditure and savings numbers from large federal energy efficiency efforts with results shown in Table 1.\textsuperscript{17} Voluntary programs appear to have energy savings on the same scale as some mandatory programs with small federal government costs, but voluntary program results are uncertain and difficult to verify. Mandatory residential appliance standards and utility demand-side management (DSM) programs both show benefits at more than twice the cost even without considering environmental costs.\textsuperscript{23}

A. Efficiency standards

Conservation activists insist that appliance efficiency regulations are needed because consumers notice an increase in purchase price but give less attention to the lower electricity payments over time. Regulations initiated in California and other states were later adopted at the federal level. Standards covering devices from washing machines to exit signs to mobile homes have had a large impact on end user efficiency. Federal appliance efficiency standards began in earnest with the sweeping 1987 National Appliance Energy Conservation Act and have been supplemented and updated frequently.\textsuperscript{24} Table 1 shows that residential savings from appliance efficiency standards in the year 2000 are estimated to be $42.32/MWh,\textsuperscript{17} less than half the retail residential electricity price of $96.42/MWh.\textsuperscript{25,26}

Building efficiency codes have developed similarly, with an indispensable role played by professional societies. In 1977 the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE)\textsuperscript{27,28} and the Council of American Building Officials (CABO)\textsuperscript{29} developed initial versions of their energy codes for commercial and residential buildings respectively. Every state had instituted a building energy code based on one of these standards before the 1992\textsuperscript{30} Energy Policy Act mandated them.\textsuperscript{31} Given the high level of technical complexity and domain

<table>
<thead>
<tr>
<th>Program</th>
<th>Energy\textsuperscript{19} Savings Quads/Year</th>
<th>$Billion/Year</th>
<th>Costs Reported</th>
<th>Cost-Effectiveness ($/MWh)\textsuperscript{20}</th>
<th>Retail Price\textsuperscript{21} ($/MWh)</th>
<th>Benefit–Cost Ratio\textsuperscript{22}</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mandatory</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appliance standards</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residential</td>
<td>0.77</td>
<td>$2.81</td>
<td>Consumer, manufacturer</td>
<td>$42.57</td>
<td>$96.42</td>
<td>2.26</td>
</tr>
<tr>
<td>Commercial</td>
<td>0.43</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Utility DSM</td>
<td>0.62</td>
<td>$1.99</td>
<td>Utility</td>
<td>$37.50</td>
<td>$79.69</td>
<td>2.13</td>
</tr>
<tr>
<td><strong>Voluntary</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EnergyStar\textsuperscript{23}</td>
<td>Less than 0.93</td>
<td>$0.06</td>
<td>Government</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>1605b Registry</td>
<td>Less than 0.41</td>
<td>$0.0004</td>
<td>Government</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>DOE Climate Challenge</td>
<td>Less than 0.81</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
expertise necessary to develop and maintain these standards, the roles of ASHRAE and CABO have been essential.

**B. Demand-side management**

In the mid-1970s, California and Wisconsin ordered utilities to work with customers to increase energy efficiency. Congress picked up DSM in the 1978 National Energy Conservation Policy Act. Utilities were expected to treat peak demand reduction as an alternative to capacity growth from an integrated resources planning (IRP) perspective. During the next decade the meaning of DSM evolved to incorporate efficiency as well as load profile management. Since utilities were compensated for their DSM programs and reported energy savings without a detailed audit, some analysts were skeptical of the reported savings, but Parfomak and Lave used *ex post* econometric analysis to verify that 99.4 percent of the reported savings were statistically observed. DSM programs have incorporated educational materials, appliance rebates, subsidized loans, customer audits, and direct installation. The education and loan components have not proven their effectiveness, but are boosted by engaging marketing materials. Rebates are popular but have questionable impact in changing consumer choices. Audits and installation with cost sharing are effective but can only reach small numbers of customers, given their complexity and labor intensity. Effective programs have drawn on the efforts of other interested parties such as product manufacturers.

Effective DSM programs are expensive and labor-intensive. Utility DSM programs grew increasingly sophisticated, effective, and costly from their conception until their peak expenditure in 2003, partially shown in Figure 1. The RFF 2000 cost estimate for avoided energy from DSM programs is $37.74/MWh which shows slightly better performance than appliance efficiency standards with benefit-cost ratios of 2.28 and 2.11, respectively. With industry restructuring, DSM expenditure declined dramatically, as shown in Figure 1. Restructuring focused on lowering price and there was less ability to hide the program expenditures from customers. Incremental energy and peak savings from efficiency efforts have generated net benefits. From Figure 1 it appears that load management expenditure had almost no payoff in energy savings and a volatile relationship with peak shaving. Peak shaving spiked just as much of the industry was preparing for restructuring, even though load management investments were on a steady decline. This might indicate that utilities were increasing accountability for coincident peak load. Some federal and state efforts have tried to stem the decline in efficiency investments with public benefit.

![Figure 1: DSM Expenditure and Savings 1989–2004](image-url)
funds such as the Low Income Weatherization Assistance Program, which may account for the increased expenditures on efficiency after 1998.

C. Energy services companies

The energy services sector was created by DSM programs. Some utilities created subsidiaries for the DSM programs while others contracted with independent companies. In 2000, 90 percent of all energy services company (ESCO) revenues were earned by subsidiaries of an energy company, as shown in Figure 2.43 Although independent ESCOs are numerous, they are not nearly as large as their subsidiary competitors.44 Initial ESCO industry revenues were from performance-based contracts,45 but have shifted toward packages of services including procurement and risk management. Throughout the entire restructuring period of the late 1990s, ESCOs have continued to grow; market revenues first hit $2 billion in 2000.43

III. State of Demand Response Technology and Policy

Customers benefit from demand response and load shifting by using less expensive energy. System benefits from economic load response should be larger than end user benefits per unit, since they include congestion relief, improved reliability, and a lower capacity requirement.

Day-ahead prices have been used in nearly all related programs and demonstrations to date, possibly to allow the end user time to plan and respond without having to invest in automated enabling technology. Even though the day-ahead price is a strong predictor of the RTP, it cannot communicate unforeseen system conditions such as unplanned outages or other emergencies. System benefits from immediate load curtailment and load shedding in contingency situations can only be garnered from active load management or immediate prices, for example, PJM’s five-minute LMPs.46

Immediate response requires automated enabling technology that acts on behalf of the end user in response to an electronic price broadcast. Providing customers with information on both real-time and day-ahead prices would allow both planning and real-time response.

A. Real-time pricing

Electric utilities, distribution companies, and other retail entities buy electricity from the wholesale market and sell it to the end user. Most of the roughly 70 utilities that offer RTPs in the U.S. developed optional programs in the mid-1990s in order to retain large industrial customers under the threat of retail competition or relocation. Other primary motivations were to lower peak consumption, to encourage overall load growth, and to comply with a mandate. These non-exclusive motivations are shown in Figure 3.47 These utilities tend to offer implicit hedges to protect valuable customers from price spikes.

When some utilities offered all their large customers the option of RTP, they did so knowing that some would pay lower average prices without making any changes. Because some utilities never expected customers to respond to the RTP, it is not surprising that only 35 percent of them offered technical assistance for RTP response, and only 49 percent provided customers a way to monitor usage regularly.48 What is surprising is that these
utilities have reported 12-33 percent reduction in participants’ coincident peak load.47

B. Economic load response

Even if a current consumer paying a fixed tariff learned the five-minute LMP values by looking at the PJM Web site,46 the price would be irrelevant since the consumer would face a fixed price. Although operational demand response programs have yet to demonstrate large enrollment and responsiveness, most market operators in the United States offer some combination of economic load response, emergency response, and ancillary service programs as shown in Table 2.

The economic load response programs within ISO-NE, NYISO, and PJM are similar. If a customer is not large enough to interact directly with the wholesale market, it must participate in demand response programs via a licensed curtailment service provider (CSP). Minimum individual or aggregated curtailment is 100 kW in PJM and ISO-NE. At low prices, load usually has the option to respond to day-ahead prices but will only be compensated for curtailment when prices are above $75/MWh in PJM or $100/MWh in ISO-NE. Reporting and metering requirements are extensive; curtailments are verified based on a weather-adjusted customer baseline usage.

Double-counting is implicit in these programs because load not only chooses not to pay for the power, but also receives a payment. The customers that do not participate benefit from lower electricity prices. Curtailment payments do not reflect systems benefit of response; they were set at an arbitrary level to jumpstart enrollment.

Even though PJM, ISO-NE, and NYISO compute day-ahead and real-time locational marginal prices (LMPs) for every bus in the system, only a subset of these are posted online in real time.46

### Table 2: Market Operator Demand Response Programs

<table>
<thead>
<tr>
<th>Market Operator</th>
<th>Economic</th>
<th>Contingency</th>
<th>Ancillary Services</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAISO</td>
<td>None</td>
<td>Voluntary load reduction, investor-owned utility curtailment</td>
<td>Non-spinning reserve, replacement reserve, supplemental energy</td>
<td>500 MW in VLRP, up to 800 MW shaved in 2005</td>
</tr>
<tr>
<td>ERCOT^54</td>
<td>None</td>
<td>Included in ancillary services</td>
<td>All ancillary services</td>
<td>2.5% of total load is registered</td>
</tr>
<tr>
<td>ISO-NE</td>
<td>Day-ahead, real-time</td>
<td>Emergency</td>
<td>Investigating stage for operating reserves</td>
<td>Up to 5% of peak demand in emergency</td>
</tr>
<tr>
<td>MISO</td>
<td>None</td>
<td>Emergency</td>
<td>None</td>
<td>–</td>
</tr>
<tr>
<td>NYISO</td>
<td>Day-ahead</td>
<td>Emergency</td>
<td>Installed capacity or special case</td>
<td>2,300 customers, $75 million in capacity revenues</td>
</tr>
<tr>
<td>PJM^55,56</td>
<td>Day-ahead, real-time</td>
<td>Emergency</td>
<td>Limited ancillary services including spinning reserve</td>
<td>6,000 commercial and industrial customers, more than 45,000 small customers^56</td>
</tr>
<tr>
<td>SPP</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>–</td>
</tr>
</tbody>
</table>
Demand response programs are settled at the aggregate zone level. This averaging prevents localized congestion from being reflected and alleviated through demand response. The Internet-based communication system used in ISO-NE to transmit the real-time zonal prices might be the most advanced system in operation. Responders in New England can receive up to $2,800 in reimbursement for compatible communications devices.

Back-up generation can be employed in these programs with proper permitting, but not if the same resource receives capacity payments. Table 3 shows the sizes of several contingency and market-based demand response programs, many of which are not operated by ISOs or RTOs. Actual curtailments are much higher in contingency programs than they are in economic response programs, possibly because involvement is sometimes binding. Back-up generation serves as a significant but not overwhelming proportion of curtailed load.

### C. Load in ancillary service

Using load as an ancillary resource is an old idea that has been developed for specific applications from voltage support to spinning reserve, to stochastic frequency control. National laboratory projects have also demonstrated the technical feasibility of using municipal pumped water and residential air conditioners to provide spinning reserve. Incorporating load as a regulation and reserve resource might become even more important if wind resources grow into a significant generation asset.

Many enacted projects fall under the category of demand response in ancillary services. Most common among these are emergency load curtailment programs instituted by investor-owned utilities. Market operators also employ load shedding under stress; in PJM an emergency responder collects either $500/MWh or the zonal LMP, whichever is higher.

### IV. Magnitude of Electric Energy Savings

Comparing the magnitude of possible savings between efficiency and demand response is important for guiding public and private investments. The comparison is difficult because must demonstrate both ability to respond and the level of response before the market operator will recognize bids. These programs have been developed and implemented quickly considering that the Federal Energy Regulatory Commission (FERC) and the North American Electric Reliability Council (NERC) regulations only began allowing for ancillary services on the demand side beginning in 2002.

Including loads as a resource became possible when national standards moved away from prescriptive standards of how ancillary services should be provided and toward performance-based standards. Regional reliability bodies and market operators can still decide whether to allow demand-side provision of ancillary services. ERCOT appears to lead the market operators in providing technical and market tools for the private sector to use in integrating load into ancillary provision. In its Load Acting as a Resource program, ERCOT will employ load for any ancillary service as long as it is enabled with the stipulated communications and control devices.

#### Table 3: Response Rates and Back-up Generation Proportion of Several Demand Response Programs

<table>
<thead>
<tr>
<th>Program Type</th>
<th>Number of Programs</th>
<th>Average Curtailable Load (MW)</th>
<th>Average Load Curtailed (MW)</th>
<th>Percent of Enrolled Load Actually Responded</th>
<th>Back-Up Generation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contingency</td>
<td>8</td>
<td>158</td>
<td>84</td>
<td>64</td>
<td>31</td>
</tr>
<tr>
<td>Market</td>
<td>10</td>
<td>204</td>
<td>21</td>
<td>17</td>
<td>12</td>
</tr>
</tbody>
</table>

April 2007, Vol. 20, Issue 3

© 2007 Published by Elsevier Inc., doi:10.1016/j.tej.2007.01.006
energy savings are most important in evaluating efficiency investments while peak load reduction is most important in evaluating demand response. Savings will be informed by exploring retrospective and prospective estimates.

A. Energy efficiency savings

An energy efficiency savings projection relies on the combination of an economic model and a policy scenario. A 1999 study that analyzed environmental energy policies over the entire United States projected electric savings of 5 percent in a moderate and 11 percent in an advanced policy scenario. A set of nine prospective efficiency savings estimates from seven studies is featured in Figure 4. The national study and five state or regional studies show variability stemming from policy assumptions, locational differences, and fundamental uncertainty. Not all of the studies make separate “achievable” and “economic” estimates, but the ones that do have lower achievable projected savings because some upgrades that would pay off cannot be implemented for practical reasons.

These nine studies project a range of 10 to 33 percent in potential energy efficiency gains from aggressive policy changes. Policy strategies included in these studies reflect efforts similar to traditional demand side management tools and have time horizons from five and 20 years.

B. Elasticity of demand

Many analyses and experiments have been undertaken in order to examine price responsiveness as well as the responsiveness to shifting demand to a lower cost hour. Some experiments are more relevant to demand response because they examine responsiveness to day-ahead hourly prices or with enabling technology. Results are highly variable, partly because responsiveness behavior is complex and highly dependent on the details of the experiment including how prices are communicated. For example, if customers are recruited into a program by being assured that they would not have to pay a higher bill than if they had not participated in the experiment, their incentives are eroded. Similarly, if they know the program will last for only a year or two, they have little incentive to replace appliances or make a capital expenditure that would pay off under a long-term program.

Price responsiveness is much greater when customers have an incentive to react by purchasing more efficient appliances and equipment; in the short run, end users can reduce usage only by forgoing consumption. A 1984 review of 34 short-run and long-run estimates found median elasticities of −0.20 and −0.90, respectively, implying that a 10 percent price increase would reduce consumption by 2 percent in the short run and 9 percent in the long run. Over the long run, these same customers can make additional choices about buying efficient appliances and equipment. Figure 5 shows the difference between short-run and long-run responsiveness.

A recent Department of Energy review published price elasticities of substitution under TOU, critical peak pricing (CPP), and day-ahead real-time price (RTP) situations. Figure 6 shows averages and ranges reported from four of these studies in residential and commercial and industrial (C&I) sectors. The range of elasticities of substitution was 0.02 to 0.27.

Figure 4: Economically Feasible and Practically Achievable Electric Savings
In the future, short-run price elasticity and elasticity of substitution will depend on the sophistication of enabling technology. Modern electronics allow customers to respond to each price change without further thought or effort by having an “energy manager” run electric hot water heaters, dishwashers, pool pumps, and air conditioners during less expensive hours.

C. Demand response savings

Projecting the savings in a switch from an average-price system to a real-time price system is complicated by the uncertainty in how customers will respond. Borenstein has projected that if all customers faced the RTP, equilibrium customer dollar savings would range from 2.0 to 13.7 percent depending on the responsiveness of demand. Table 4 shows the projected savings when different fractions of load face the RTP and the demand elasticity is $-0.1$. Coincident peak load reductions are large, implying that RTPs would indeed be an effective means of addressing peak demand problems.

Overall energy consumption actually increases under this model because customers can increase usage when prices are low. An increase in energy consumption or profile-dependent pollution under RTP is a real concern. One effect that this model does not address is that responsive customers who have greater control over when they use electricity would also have greater control over whether they use electricity. For example, the Carrier ComfortChoice thermostats that have been used to demonstrate spinning reserve from load reductions also allow customers to specify timed usage. A homeowner can leave the air conditioner off all day while she is at work and have it turn on in time to return to a cool house; she can also control the device over the Web if she forgets to turn it off before a vacation.

One question to ask is whether most of the savings from RTPs could be gained from applying the much simpler time-of-use (TOU) rates. Borenstein has projected that when switching from flat-rate tariffs, total economic surplus increases with TOU rates are only 8 to 29 percent of the surplus increases with RTPs, as shown in Figure 7. The surplus increase is expressed as a percentage of customer baseline expenses. The three TOU rate schedules represent progressively more detailed price granularity. This indicates that if end users really can be responsive in real time, then the savings from

Table 4: Equilibrium Savings in Switching from Average Price to RTP, Elasticity $-0.1$^74

<table>
<thead>
<tr>
<th>Participating Load (%)</th>
<th>Customer Bills ($)</th>
<th>Energy Consumption (MWh)</th>
<th>Peak Power (MW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>33.3%</td>
<td>3.51%</td>
<td>-0.53%</td>
<td>14.0%</td>
</tr>
<tr>
<td>66.7%</td>
<td>5.25%</td>
<td>-0.92%</td>
<td>20.3%</td>
</tr>
<tr>
<td>99.9%</td>
<td>6.52%</td>
<td>-1.23%</td>
<td>24.5%</td>
</tr>
</tbody>
</table>

---

the most accurate price signals are substantially greater than those from TOU.

V. Barriers to Electricity Market Efficiency

A frustration to policy makers is the continued inaction to reap the savings when an investment in energy efficiency would have a high return. Some failures to invest in efficiency appear irrational from the engineering economic analysis but make sense when hidden costs are included. Other parts are viewed as market failures. Either way, advocates site barriers to realizing efficiency investments as reasons to enact correcting policy.

Most of the recognized barriers in adopting energy efficiency technology will also inhibit the adoption of demand response technology and strategies; some of the same impedances have already been noted.78

A. Hidden costs to efficiency

Few customers can, or have the time to, calculate the return to energy efficiency investments. A more subtle barrier to implementing efficiency programs might be a limited range of features in the efficient models.79 The EnergyStar program informs consumers about which appliances are efficient with an accessible labeling system at very low cost to the manufacturer or federal government, although the resulting benefits are difficult to quantify.17

High-level macroeconomic models that attempt to evaluate economically efficient outcomes are not detailed enough to capture hidden costs at the technology level where they occur. Accounting for the engineering economics of current physical capital and investment costs is becoming a more important part of policy modeling. A proxy for hidden costs is included in the National Energy Modeling System by introducing technology adoption rates and hurdle costs. Models that incorporate these hidden costs explicitly tend to have outcomes with lower energy efficiency.80

B. Non-cost barriers to efficiency

Lack of consumer knowledge about energy efficiency and related costs can be seen as a market failure. End users may not be able to afford the more efficient appliance or might be financing the purchase with a credit card. Many efficiency investments that are attractive at social rates of return of 2 to 5 percent are unattractive at credit card interest rates of 18 percent or more. Some other situations lead to suboptimal efficiency investments. When different budgets are used for technology investments and for energy costs, the incentive is to decrease up-front costs even at the expense of long-term gains. This situation is acute in a landlord-tenant situation where a landlord buys the least expensive, inefficient air conditioning equipment but the tenant will have to pay the electric bills.81 A similar situation can occur even within one firm with a putative common bottom line; the purchasing department might try to minimize the cost of procuring...
lighting fixtures without considering the long-term electric costs that will be paid by facilities management. Still another situation arises when firms have capital budgets with hard limits; such firms may refuse to buy efficient products regardless of payback. At any rate, once technology is installed, the energy efficiency decision is unlikely to be undone until the end of equipment lifetime; the only changes that can be made until the equipment is replaced are laborious behavioral and usage changes.

VI. Outlook

A. Load as a driving force

When loads are subjected to RTPs, customers will react to the prices and may invest in automated demand response with the help of a load aggregator. Internalizing the externalities from limiting emissions of pollutants and greenhouse gases will increase the average cost of power; it is unclear what effect it will have on the relative cost of peaking power. However, no reactions can occur unless customers know the price in real time.

Initial load response will reflect the easiest and cheapest ways of reducing expense. Figure 8 shows the response strategies used by Niagara Mohawk Power Company’s large customers under mandatory RTP billing. Among firms that reported shifting load, 47 percent said they would shift to the next day, 18 percent to the following day, and only 35 percent to another time of day. Evidently time of day is more important than actual day in consuming electricity, possibly because of scheduled shifts and operations. Large customers might find it too expensive and disruptive to flatten their load profiles, even if they are willing to make some changes.

Among firms that reported forgoing load, 65 percent said it had minimal or no impact on facility operation, 20 percent reported significant inconvenience or discomfort, 9 percent had to adjust business operations, and 6 percent reported not knowing. If many firms can respond to high prices without impacting their missions, then some of the benefits of demand response can be achieved without significant drawback.

Although regulators might be hesitant to impose RTPs for fear of end user pushback, in Mohawk only 15 percent of customers were dissatisfied with a switch to RTPs from TOU, even though 54 percent reported that they did not respond in real time. Some customers, especially governmental and educational facilities, report that they have responded to system emergencies not because prices were high but rather because it was a civic duty. The only customer who would protest the RTP would be one who refused to change her usage and who used more power during the peak hours and so was free-riding on customers who used more power during the off-peak hours.

B. Opportunity for energy services companies

A study of 1,379 recent ESCO projects shows that these companies are cost-effectively upgrading the electric efficiency for their clients. When ESCOs have upgraded lighting equipment, they have delivered median energy savings of 47 percent on lighting equipment. When ESCOs have performed services beyond lighting, they
have delivered median savings of 23 percent\textsuperscript{87} from the entire electric bill. Figure 9 shows the percentage of these projects that have made improvements of various types. Traditionally inefficient systems such as lighting and heating ventilating and air conditioning (HVAC) are often addressed, but a significant portion of projects involve “other” services as well. These other services can be backup fuel choices, training, or rate analysis.

Energy services are a growing market that would find more opportunities for growth if many more customers are subjected to RTPs. Although most of the customer base for ESCOs is in publicly funded facilities, 26 percent of revenues are from the private sector, especially office space and industrial facilities.\textsuperscript{43} Demand response can be added to the portfolio of packaged services that ESCOs offer. Some market operators appear to value ESCOs as intermediaries between the load and the marketplace, but not all market operators offer demand response programs.\textsuperscript{15,16}

Market rule changes and additional communication services might be necessary for ESCOs to offer demand response and these needs should be communicated to the market operators.

C. Market operator responsibility

A signal that enables an automated response is much more valuable than prices that customers must seek out on a Web page or receive via phone, email, or fax. Continuously checking day-ahead or real-time LMPs is too laborious for many loads.\textsuperscript{85} The most useful communications formats might be RSS Web feed, text messaging, or paging; market operators can probably determine the most useful medium for broadcast simply by asking curtailment service providers what they can use. The Internet-based economic communication system within ISO-NE is a model for other markets to follow. Similarly, ERCOT has done a very good job of opening ancillary markets to market and making the necessary communications and control devices available.

D. National standards

In general, choices of demand response technology, communication, and contractual structures need not be decided by FERC or NERC. The role of regulators and standards bodies is to open markets to competition and participation for all generators and loads. Stipulating that large end users must face RTPs is a prerequisite to making demand response possible without subsidy.\textsuperscript{88} Although FERC standards make it possible for demand to have equal opportunity for energy and ancillary market participation, the stipulation that large end users must face RTPs has not happened nationally and may require federal or state legislation.

A form of time-of-use pricing happens in deregulated markets when a broker buys electricity in the wholesale market for customers. The price that the broker can offer depends on the time profile of company usage. The broker can show a customer how much the total electricity bill will decline by shifting some demand to off-peak hours. Similarly, the broker can contact customers to tell them that the current wholesale price is very high or very low and that they can lower their total bill by lowering consumption so that the electricity need not be purchased or can be sold back into the market or additional electricity can be purchased at the low price.

Small customers may not offer enough system benefit to warrant

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure9.png}
\caption{Percent of ESCO Projects That Employed Various Cost-Saving Strategies\textsuperscript{43}}
\end{figure}
the expense of time-of-use or real-time metering. Using the observed variability of whole prices, the expense of a smart meter, the consumption level of a consumer, and the likely response to higher prices, it is straightforward to estimate whether installing a smart meter will benefit the customer. Aggregators have already organized customers into large loads to realize savings.\textsuperscript{56} Eventually aggregators will organize even residential customers if there is profit in it. Requiring large loads to face RTPs does not mean that they cannot get a flat rate contract; a broker would be willing to offer any contract that the customer wants, at a suitable price. Similar implied hedges have already been observed in RTP tariffs.\textsuperscript{47} Customers who purchase level prices still have incentives to reduce peak load because the level price be set higher for a customer with high on-peak consumption.

VII. Challenges and Opportunities

We emphasize customer response as a neglected way of solving electricity industry problems. Historically, providers have focused on supply, assuming that consumers are unwilling or unable to modify their consumption. Contrary to these expectations, customers respond to higher prices that they expect to continue by purchasing more efficient appliances and other efficiency measures. When there are power shortages, customers have shown that they will respond to pleas to reduce demand. Large industrial and commercial customers currently respond to time-of-use and real-time pricing. With the addition of an electronic energy manager, small consumers could respond in real time to price fluctuations. This customer response has the potential to lower costs by eliminating the most expensive peaking generators, as well providing ancillary serves on the demand side and virtually eliminating blackouts.

Restructuring electricity markets has exacerbated an irritant by paying the market clearing price to all generators, providing an almost irresistible incentive to manipulate the market. One way to lower demand is to have consumers understand the implications of their purchases of appliances and other devices that use energy. In many cases, consumers purchase inefficient air conditioners, hot water heaters, and other devices, although paying a bit more for an efficient appliance would save money over time. Government programs attempt to deal with the situation by requiring appliances to have prominent efficiency labels and by setting minimum standards. While much has been accomplished here, much remains to be done in situations where the person paying the electricity bill does not select the appliance or the person making the purchase does not have the money to buy the more efficient appliance.

Another important way to achieve savings is to allow end users to stop buying additional kWh when the RTP exceeds the price they are willing to pay. Just as consumers have learned to respond to the volatile prices of gasoline, fruits and vegetables, and other commodities, so they can learn to respond to electricity prices. The largest difference is that customers purchase electricity every hour of the year and therefore need automated devices to react to changing prices without spending all their time looking up prices and making decisions.

While some policymakers and utilities fear that consumers will protest RTPs, experience has found few unhappy customers. Even if they do not change their usage patterns, most customers would find no change in their total bills, since they already pay the average of all high- and low-price hours. Those customers who do choose to react to the high-priced hours would lower their own bills, and even lower the bills of unresponsive customers because peak prices would fall.
A service provider or market operator already has sufficient information to inform individual consumers as to the real-time LMP of electricity. The principal barrier to RTP is the installation of an hourly meter. As the current stock of energy meters are replaced, they should be replaced with real-time meters. Smart meters do not necessarily have to be monitored in real-time, if they record hourly consumption data. Additional communication expense is incurred if an LSE is to monitor real-time usage and provide the customer with this information. Some retailers already find it worth their while to install communications with their meters so that they do not have to pay the labor costs of meter readers. Customers must decide for themselves whether to invest in automated devices or ESCO services that would allow them to react to the RTP.

Demand response will become more important as electricity prices rise due to fuel price increases, the need for new generation and distribution, and some of the price increases that have come from unfreezing prices after deregulation. Investment in expensive new capacity can be obviated by demand response and market clearing prices can be lowered. As wind power realizes large-scale deployment, the ability of load to shift power use to coordinate with availability will become more valuable and essential. When carbon constraints are included in electricity prices, reducing end user cost will become more important. Customer ability to respond and adapt to these additional costs and system pressures will be greater, with more accurate price signals and greater load response. Demand response capability can be part of an overall package of services and greater controls offered by ESCOs. The adaptability that ESCOs have exhibited through deregulation will be invaluable when taking on the additional challenge of making demand response available to consumers.

Endnotes:

1. The term “market operators” will be used to refer to both ISOs and RTOs. When referring to a single state or multi-state entity, the acronym ISO or RTO will be used as appropriate.

2. California ISO is the exception, with a $400/MWh soft cap on energy and ancillary service bids; see reference in note 6. Generators may bid above a soft price cap and will be paid as bid; other generators will receive payment only as high as the cap. The neighboring Western Electricity Coordinating Council (WECC) has the same price caps although WECC is not a market operator.


8. Retail prices are higher than wholesale prices because of transmission, local distribution, billing, and local service costs.


18. Costs and benefits result from all programs or standards up until the year 2000; the numbers have been annualized so that the costs and benefits can be viewed over a one-year slice of time. These and all dollar values in this article are updated to $2006 using Bureau of Labor Statistics inflation data; see, Inflation Calculator, Bureau of Labor Statistics, available at http://www.bls.gov/.

19. Energy savings are reported in quadrillion BTUs of source energy.

20. Cost effectiveness numbers are reported assuming that all energy is converted to electric energy. A conversion factor of 11660 BTU source/kWh electric corresponding to a conversion efficiency of about 29 percent was used.

21. Year 2000 residential and average retail prices are reported for comparison with residential appliance standards and utility DSM programs respectively; see note 25.

22. Benefit-cost ratio compares benefits accrued to the end user to costs reported. Environmental benefits and costs to unlisted parties are not considered.

23. EnergyStar cost and savings numbers are reported for the year 2001; all other program numbers are reported for the year 2000.


26. Original RFF and EIA numbers were both converted to 2006 dollars. The EIA number refers only to retail residential sales; commercial and industrial rates were $76.7/MWh and $47.9/MWh, respectively.

27. The ASHRAE 90.1 series with its periodic updates has been widely adopted for commercial and high-rise residential facilities. Forty states have adopted a version of this code; see note 28.


29. The CABO developed the original Model Energy Code (MEC) for residences, which is now the International Energy Conservation Code (IECC). Some version of this code is enacted in 40 states. The list of noncompliant states is not identical between commercial and residential sectors; see note 28.

30. Federal standards were again updated in the 2005 Energy Policy Act to reflect the most recent versions of these codes, ASHRAE 90.1-1999 and IECC.


36. Spending trends are from the EIA as estimated from numbers reported by utilities.

37. See Table 1.

38. Total utility DSM expenditure includes indirect costs as well as efficiency and load management costs. Indirect costs represent between 8.5 percent and 17.7 percent of total expenditure in this time period.
39. Only total expenditure data were available prior to 1993 because the EIA did not collect the more complete data before that year.
42. “Incremental” savings are attributable to expenditures in the current year, not from previous years. The EIA also reports “annual” numbers that represent current year savings from all previous investments.
44. “Equipment Manufacturers” generally make building equipment and controls; “Other Energy Companies” can be fuel producers, pipeline owners, etc.
45. A performance-based contract is an arrangement in which an energy services company will install efficiency upgrades for a client. The client and the energy company then share the savings accrued from the lower energy bills.
48. “Regular” means real-time energy consumption or consumption information from the previous day.
49. All of the members of the ISO/ RTO Council that are in the United States are examined here. California ISO (CAISO), Electric Reliability Council of Texas (ERCOT), ISO New England (ISO-NE), Midwest ISO (MISO), New York ISO (NYISO), Pennsylvania–New Jersey–Maryland (PJM) RTO, and Southwest Power Pool (SPP).
50. Market operators that do not offer economic response programs state that they allow price response via a third-party intermediary, but do not support such response with electronic price broadcasts.
51. All ancillary services here require that load have automated response to identical signals given to generators and demonstrate their ability to respond.
52. CAISO used to offer more programs but has eliminated them as investor-owned utility distribution companies (UDCs) have increased their own load curtailment programs.
53. Supplemental energy is a near real-time response.
58. Percentages are an average of percentages for individual programs, not a percentage of totals from all programs.
65. These numbers represent yearly savings after a 10-year time horizon based on the Energy Information
Administration (EIA) base case projection.


67. Source reports higher possible gains from technically feasible but economically infeasible options; only “economic” or “achievable” results are examined here.


69. Price elasticity of substitution is a measure of load shifting in this context, generally measured between on-peak and off-peak hours. There is no standard definition of peak hours.


71. Appendix C of the source in note 70 contains a review of elasticity studies and their relationship to demand response. Elasticity numbers reported here are obtained from this source.

72. The short-run numbers were recently updated in another review of 36 estimates with a median of −0.28.

73. This simulation used California market loads and a mix of three generator types.


75. The elasticity reflecting these estimates ranges from −0.025 to −0.5. It also reflects scenarios in which demand is more, less, or equally responsive during peak times.

76. For example, if peak generation is natural gas and baseload is coal, a flat load profile would increase emissions of sulfur, particulates, mercury, and other pollutants which are much higher from coal generation than from gas generation; see note 82.

77. The sum of consumer and producer surplus is the total surplus. Under all TOU and RTP pricing structures examined, the total surplus increased compared to the total surplus under flat-rate pricing. To scale the total surplus increase numbers, they are expressed as a percentage of the entire consumer electric bills under the flat-rate scenario.


84. Percentages do not sum to 100% because response categories are non-exclusive. Sectors had response numbers N = 10, 26, 10, 23, and 8 from top to bottom.


86. The 50% confidence interval is 37% to 56%.

87. The 50% confidence interval is 17% to 32%.
