

# Time-Varying Retail Electricity Prices: Theory and Practice

by

Severin Borenstein

**Suggested Running Head:** “Time-Varying Electricity Prices”

**Author's Blurb:** Severin Borenstein is E.T. Grether Professor of Business and Public Policy at the Haas School of Business, University of California, Berkeley. He is also Director of the University of California Energy Institute and a Research Associate of the National Bureau of Economic Research. Steve Braithwait, Jim Bushnell, Ahmad Faruqi, Jim Griffin, Stephen Holland, Michael Jaske, Jennifer Kaiser, Karen Notsund, Steve Puller, Art Rosenfeld, Larry Ruff, Celeste Saravia, Frank Wolak and Catherine Wolfram provided valuable comments. Jennifer Kaiser contributed excellent research assistance. An earlier version of this paper appeared as part of Borenstein, Jaske and Rosenfeld, "Dynamic Pricing, Advanced Metering and Demand Response in Electricity Markets," an unpublished study sponsored by the Hewlitt Foundation and the Energy Foundation ([www.ef.org](http://www.ef.org)), also released as Center for the Study of Energy Markets Working Paper #105, University of California Energy Institute. Financial support from The Hewlitt Foundation and the Energy Foundation is gratefully acknowledged.

Electricity is not economically storable and production is subject to rigid short-term capacity constraints. Since demand is highly variable, this means there will be times when there is plenty of capacity and the only incremental costs of producing electricity will be fuel and some operating and maintenance (O&M) costs. At other times, the capacity constraint will be binding, causing the incremental cost to increase greatly, and wholesale market prices to rise. Supply constraints are even more likely if sellers are able to exercise market power, exacerbating the volatility of wholesale prices.

The result of this structure is that the wholesale price of electricity, reflecting the supply/demand interaction, varies constantly. In most markets, the wholesale price changes every half-hour or hour. The end-use customer, however, sees the retail price, which typically is constant for months at a time. The retail price does not reflect the hour-by-hour variation in the underlying wholesale cost of electricity. A number of programs have been implemented or proposed to make the economic incentives of customers more accurately reflect the time-varying wholesale cost of electricity. Opponents have expressed concern that these programs expose customers to too much price volatility. Still, time-varying retail prices hold the key to mitigating price volatility in wholesale electricity spot markets.

In the first section of this paper, I present the theoretical argument in support of time-varying retail prices and explain the societal loss from using flat rate retail pricing. I also examine the role that time-varying retail prices can play in reducing the exercise of market power by sellers. In section II, I discuss the variety of programs used implementing time-varying retail prices and discuss their effectiveness in giving customers economically efficient incentives. In section III, I focus primarily on real-time retail pricing (RTP) and examine a number of issues and complaints that frequently arise in implementation of RTP. Section IV presents the results of some simulations of the effect that RTP would have in the long run on prices, consumption and investment, suggesting that the benefits from RTP for large customers, at least, are likely to far exceed the costs. Section V concludes.

## **I. The Economic Efficiency of Time-Varying Retail Price**

For most of the 20th century, electricity was sold in regulated environments in which the retail price did not vary based on the time it was used. Customers faced a constant

price for electricity regardless of the supply/demand balance in the grid. Discussions of changes to greater variation in retail price have usually focused on who, among customers, would win or lose from such a change. While the distributional impact among customers is certainly important, time-varying electricity prices are also very likely to affect the total cost of the electricity and, in the short run, the allocation of the cost between customers and deregulated producers. In this section, I examine the economic impact of moving to a system in which prices more accurately reflect the supply/demand balance.

### 1. *Efficient Time-Varying Electricity Pricing*

For illustration, I will assume that there are only two levels of demand, peak and off-peak, as shown in Figure 1, and that all producers have the same cost of production up,  $c$  to their capacity. I'll begin by illustrating how this market would operate if different prices were charged during peak and off-peak times and no producer were able to exercise market power. I return to the market power issue below.

[FIGURE 1 ABOUT HERE]

If the market were competitive and total installed capacity were  $K$ , the market supply curve would be flat at  $c$  out to  $K$  and then vertical. It is easy to see that the prices in the market would be  $P_p$  and  $P_{op}$ . If any producer tried to charge a price above  $P_{op}$  during off-peak, it would be unable to sell its power since there is idle capacity waiting to jump in if price is above the producers' marginal cost. During peak times, no producer would be willing to sell at a price below  $P_p$ , because  $P_p$  clears the market, *i.e.*, since there is no idle capacity, any producer can sell all of its output at that price, so it would have no incentive to charge less. If any producer tried to charge more than  $P_p$ , it would find that its unit sales would decline and it would be unable to earn enough on the power it sold to justify selling less than its full capacity, as it could at  $P_p$ . This follows from the assumption that no firm can profitably exercise market power.

Now consider the outcome if, due to either technological or legal constraints, the firms charged the same price for peak and off-peak demand. If the firms were still to break even overall, the price would lie somewhere between the peak and off-peak price from the previous example, a price we'll call  $\bar{P}$ . This is illustrated in Figure 2. For the off-

peak demand, this would raise the price and inefficiently discourage off-peak consumption, because the price would exceed the true marginal cost of production. Some consumption that would produce value greater than the incremental cost of production would not take place. This is illustrated by the vertically shaded “deadweight-loss” area.

[FIGURE 2 ABOUT HERE]

A single-price restriction would result in a price for peak demand that is below  $P_p$ . This would increase the total quantity demanded above the current market capacity. Absent additional capacity, this would cause a shortage, and would require some sort of rationing, using either economic incentives or some alternate approach. Putting aside, for the moment, the disruption of blackouts, such rationing doesn’t even necessarily deliver power to the most valuable use. A use with a value just barely above  $\bar{P}$  would be as likely to receive power as a use with a much higher value.<sup>1</sup>

In reality, of course, the excess demand at peak times is not allowed to cause blackouts. Instead, capacity is expanded to meet the high demand that results at peak times. The question then is whether this is a good use of resources. The answer is almost certainly no, as is explained in the next subsection.

## 2. *Efficient Capacity Investment*

To analyze the efficiency of capacity investment with and without time-varying prices, we return to Figure 1 to illustrate capacity investment with time-varying prices. It is clear in this situation that additional capacity has no value for the off-peak period. Off-peak consumption does not even utilize all of the currently available capacity. Additional capacity does have value in the peak period because the marginal value of power to customers is greater than the marginal cost that would have to be expended after an additional unit of capacity was built. To be exact, the value of an additional unit of capacity is  $P_p - MC$ , which is  $\lambda_p$  in the illustration.  $\lambda_p$  is the peak-period shadow value of marginal capacity at the current level,  $K$ . The off-peak shadow value of marginal capacity is  $\lambda_{op} = 0$ .

The per-“day” (where a day includes one peak and one off-peak period) cost of one additional unit of capacity is the fixed operation and maintenance cost plus depreciation and the opportunity cost of the capital investment, *i.e.*, the foregone interest on the funds

used for this investment.<sup>2</sup> I will call this full fixed capacity cost  $r$  per unit of capacity per day. The efficient criterion for capacity choice is to expand capacity so long as the sum of the capacity shadow values for all periods is greater than the capacity cost,  $\sum \lambda > r$ , and to stop expanding capacity at the point that  $\sum \lambda = r$ . Note that we take the *sum* of the  $\lambda$ s because the peak and off-peak operation are non-competing uses of the same capacity.

Luckily, this is also the criterion that will determine the competitive level of capacity since each price-taking owner of capacity will receive  $\lambda_p$  above its operating costs in peak periods and  $\lambda_{op}$  above operating costs in off-peak periods. Thus, a firm will have an incentive to expand capacity so long as the sum of the  $\lambda$ s is greater than the cost of expanding capacity.<sup>3</sup>

Now, let's return to the issue of capacity investment under a single-price retail system. At price  $\bar{P}$ , the off-peak demand is well below capacity so the shadow value of capacity for off-peak demand remains zero. In order to meet the peak demand at  $\bar{P}$ , however, we have built additional capacity,  $\Delta K$ . But, it is not efficient to build that capacity: the net value of the additional power produced by the capacity is less than the cost of the capacity. To be concrete, building the extra  $\Delta K$  of capacity creates deadweight loss equal to the shaded triangle in Figure 3 because the net value of this additional capacity is less than the cost of holding the capacity, which is  $r \cdot \Delta K$ .

[FIGURE 3 ABOUT HERE]

In the real world, this inefficiency shows up in the form of excess capacity that is underutilized, but still must be built in order to accommodate the peak demand. The value customers get out of this capacity is not great enough to justify the capital investment. With time-varying pricing of electricity, this excess capacity is not necessary because higher prices at peak times encourage customers to consume less at those times, either by shifting peak consumption to off-peak or by simply reducing consumption at peak times.

### *3. Time-Varying Prices and Market Power*

Thus far, the analysis has considered only the case of a competitive wholesale electricity market. When producers are able to exercise market power, however, the benefit of instituting time-varying prices is greater. In any market, a seller or group of sellers

exercises market power by raising price above the competitive level. The financial attractiveness of this action depends on the tradeoff of higher prices on the quantity the firm still sells versus lost sales due to the increased price. It is clear that the payoff to exercising market power is greater if raising prices has a smaller impact on sales.

When the retail price of electricity does not vary over time, a wholesale seller's attempt to exercise market power and raise wholesale prices has no short-run impact on quantity since end-use customers do not see a change in the retail price they face. This makes it much more profitable for the wholesale seller to exercise market power. With time-varying prices that reflect changes in the wholesale price, an attempt to raise wholesale prices will impact retail prices and thus reduce the quantity of power that customers demand. This customer response reduces the profitability of raising wholesale prices and, thus, discourages the exercise of market power.

To be a bit more concrete, consider a hot summer day when the system is stretched to near its limit, and consider the incentives of a wholesale seller in the market that owns 5% of the system capacity. On such a day, the wholesale seller knows that if it withdraws 1% of the system capacity from production, it will have a significant effect on the wholesale price. If, however, the retail price also increased when the wholesale price rose, then customers would get a signal that they should scale back usage. The resulting reduction in demand means that the withdrawal of production capacity would have less impact on the wholesale price and, thus, would be less profitable for the seller. In contrast, if retail prices are not linked to the wholesale price level, then there is no demand response when the seller withdraws capacity and the wholesale price is more likely to increase dramatically.

Without time-varying retail prices, the combination of supply-demand mismatches and the ability of sellers to exercise market power at peak times creates a relationship between price and system load that looks like a hockey stick laid on its side. Figure 4 shows a price/load scatterplot for California during June 2000 and a polynomial curve fitted to the points. The hockey-stick relationship is a fairly constant price over a wide range of outputs and then steeply upward-sloping price as demand grows closer to capacity. Time-varying prices would reduce the frequency and degree of price spikes during periods of high system load.

[FIGURE 4 ABOUT HERE]

## II. Time-Varying Prices in Practice

While the simple model presented in the previous section makes clear the value of time-varying prices that send accurate price signals, the actual details of implementation are, not surprisingly, more complex. In the model, there were only two time periods, and both demand and supply in each period were known with certainty. In reality, the supply/demand balance changes continuously and there can be a great deal of uncertainty about supply and demand in advance of any given period. This raises the two fundamental issues in designing time-varying retail electricity prices:

- (1) *Granularity of Prices*: the frequency with which retail prices change within the day or week, and
- (2) *Timeliness of Prices*: the time lag between when a price is set and when it is actually effective.

Flat retail rates are obviously at one extreme: No Granularity – there is a single price for day and night, weekdays and weekends – and price setting is not timely – the price is set months before some of the hours to which it is applied. At the opposite hypothetical extreme would be a real-time pricing program in which prices change every minute and are announced only at the minute in which they are applied. This hypothetical would be the ideal in terms of economic efficiency in the electricity market, but customers are likely to want more certainty about prices, an issue addressed below.

The granularity and timeliness issues in retail electricity pricing are distinct, but they are closely related and interact in somewhat subtle ways. The granularity of prices will affect the accuracy of price signals only to the extent that the *expected* cost of electricity differs across time intervals at the time the prices are set for those intervals. For instance, if prices were determined a month in advance, the value of having prices change every ten minutes rather than hourly would be negligible, because the *expected* costs for adjacent ten-minute periods barely differ one month ahead. In contrast, if prices were determined in real time, at the start of each 10 minute period, they would incorporate much more



information about idiosyncratic supply or demand shocks. Thus, there is generally greater value of fine granularity in prices when price setting is more timely.

In debates over real-time retail pricing (RTP), there is often a focus on granularity and the fact that prices change every hour in most cases. Timeliness is, in fact, at least as important. Hourly price changes would not be much of an improvement over just two or three daily price blocks per day if prices were set a month or more in advance. When timeliness is discussed as part of RTP, the issue is whether prices should be set a day ahead or an hour (or less) ahead. This choice turns out to have a large effect on the efficiency of RTP, as I discuss later.

A retail system in which prices are set in advance may at first seem to be the retail equivalent of the day-ahead and long-term contract transactions that take place in wholesale electricity markets, but there is a critical difference: The retail programs offer these prices as “requirements contracts,” meaning that at the time of delivery, the customer may choose to buy as much or as little as desired at the posted price.<sup>4</sup> If a wholesale electricity buyer purchases in advance, the purchase contract specifies both the price and the quantity. Deviations from that quantity in the wholesale buyer’s actual consumption are then priced at the spot market price in the hour in which they take place.

With this foundation, we can analyze the programs that have been designed for implementing time-varying retail prices. The possibilities range from slight augmentations on flat pricing to more radical change that would make retail electricity markets more closely resemble the wholesale electricity markets.<sup>5</sup>

### *1. Real-Time Pricing*

Depending on one’s view, either the most natural or the most extreme approach to time-varying prices is real-time retail pricing of electricity (RTP). RTP describes a system that has a very high degree of both price granularity and price timeliness. This approach is most consistent with other industries that have highly volatile wholesale prices – such as fresh fish, produce, gasoline, or computer chips – where retail prices adjust very quickly to reflect changes in the wholesale price of the good. In most designs, retail electricity prices under RTP change hourly. For each hour, say 4-5 p.m. on June 21, the price may differ

from the price at any other hour, such as 3-4 p.m. on June 21 or 4-5 p.m. on June 22.

RTP prices are typically set either “day-ahead” or “real-time.” In the day-ahead formulation, the retail provider announces all 24 hourly prices for a given day at one time on the prior day. In the real-time approach, the retail provider announces prices on a rolling basis, typically with the price for each hour determined between 15 and 90 minutes prior to the beginning of that hour. In the largest RTP program in the US, Georgia Power allows customers that sign up for RTP (all of its RTP customers are large industrial or commercial firms) to choose whether to be billed based on day-ahead or real-time announced prices. Most have chosen day-ahead.

In terms of economic incentives and efficiency, RTP using real-time announced prices offers the greatest value. The marginal incentives of consumers facing such prices most closely reflect the actual supply/demand situation in the market. Despite that attraction, most RTP implementations have used day-ahead announced prices for the hourly price that RTP customers pay. In all programs, these have been requirements contracts, so the customer can buy any quantity desired at that day-ahead price.

If RTP based on real-time announced prices is the “gold standard” in terms of economic efficiency, how much is being lost by instead using day-ahead prices as the basis for an RTP program? Part of answering that question is determining how much of the real-time price variation is captured in day-ahead prices. I’ve done this for two electricity markets that posted prices in both a day-ahead market and a real-time balancing market, California and PJM (the Pennsylvania, New Jersey, and Maryland market). On average, the day-ahead price captures 45%-50% of the variation in the real-time balancing price over a year or longer period.<sup>6</sup>

That figure alone, however, does not answer the question. Imprecise retail pricing has economic consequences only if buyers would respond to the additional information offered by more precise pricing. For instance, if the only way that an industrial customer could respond to high prices is to cancel a shift of workers, and if this decision must be made a day in advance, then there would be no efficiency loss from basing RTP on day-ahead prices, even though those price reflect the real-time supply/demand situation with some degree of noise. That is, one must ask not only how large the day-ahead forecast error is,

but also how much customers would alter their behavior if they received better information. Assuming, however, that there is some way for customers to respond, at least marginally, to changes in real-time RTP prices, retail prices that more closely reflect the real-time wholesale price will improve efficiency.

RTP, whether based on a day-ahead or real-time wholesale price – does not mean that customers must buy all of their power at the RTP price. Hedging – purchasing some power through a long-term contract, before a period of system stress is evident – allows customers to stabilize their overall bill while still facing the RTP price for incremental consumption, as I discuss in detail later.

## *2. Time-of-Use (TOU) Pricing*

While RTP has not been widely accepted or implemented, time-of-use (TOU) pricing has been used extensively in the U.S. for large industrial and commercial customers. Under TOU, the retail price varies in a preset way within certain blocks of time. For instance, a typical TOU pricing plan for weekdays during the summer charges 5.62¢ per kilowatt-hour (kWh) from 9:30 p.m. to 8:30am, 10.29¢ per kWh for 8:30am to noon and 6 p.m. to 9:30 p.m., and 23.26¢ per kWh for noon to 6 p.m.<sup>7</sup> In most cases, the weekend and holiday rates are equal to the off-peak weekday rate. The rates for each time block (usually called peak, shoulder, and off-peak) are adjusted infrequently, only two or three times per year in most cases. Price is the same at a given time of day (on a weekday) throughout the month or season for which the prices are set. TOU retail pricing lacks both the granularity and the timeliness of RTP.

The lack of timeliness of TOU prices means that they cannot capture any of the shorter-term variation in supply/demand balance. In addition, TOU programs don't reflect *expected* wholesale market variation very well, due to the lack of price granularity. The standard 3-block pricing format means that the 4-5 p.m. weekday price is the same as the noon-1 p.m. weekday price, because they are in the same block, despite the fact that the expected wholesale market price is significantly different for these hours. Furthermore, because TOU programs typically reset prices for each block only two or three times per year, the same peak, off-peak, and shoulder prices may apply from May to October, even

though the average demands and costs change in a quite predictable way during that time.

These attributes mean that TOU price variation will reflect little of the true variation in the wholesale market. Empirically, one can show this by looking at a given time period and asking how much of the wholesale price variation would be reflected in TOU prices, assuming that the TOU price were set optimally to maximize their relationship to wholesale market prices. For California, even setting TOU prices after the fact to reflect the actual average price in peak, shoulder, and off-peak periods, the TOU rates would have reflected only about 6-13% of the real-time wholesale price variation on average.<sup>8</sup> And these numbers assume that the agency setting the rates can forecast the average price in each period as accurately before the period as one can do after the fact, which is virtually impossible.

Put differently, TOU prices, while giving greater advance notice of prices and offering less price volatility, do quite poorly in reflecting variation in wholesale prices. As before, the cost of this loss of information will depend very much on how customers would react if they were given the finer information. For instance, if a factory can react to price changes only by making long-term adjustments, such as changing worker shift schedules that can be altered only semi-annually, then the information in TOU prices may be as much as the factory can use. In that case, no price-responsiveness is being sacrificed in using TOU prices instead of RTP. On the other hand, if the customer can make such adjustments more frequently, such as weekly or monthly, or can adjust quickly to idiosyncratic supply/demand information, such as by adjusting air conditioning settings and lighting when the system is strained, TOU rates won't yield such adjustments, because they fail to signal the short-term variation in the supply/demand balance.

Technology plays a role in this tradeoff. Until recently, the cost of TOU metering was substantially less than real-time metering and the ability to send real-time price information to customers was limited. Technology changes of the last decade have virtually eliminated these issues. Advances in technology also have enhanced, and continue to enhance, the customer's ability to respond to real-time price changes. Responding to frequent retail price changes does not now require human intervention. Instead, the real-time price is sent electronically to a computer that is programmed to respond. If the price goes

above 15¢ per kWh, for instance, the computer might automatically reset air conditioning from 72 to 74 degrees. A computer could also automatically reset lighting and reschedule energy-intensive activities that are time adjustable, such as running a pool pump. Thus, historical measures of firms' abilities to respond to real-time price changes are likely to significantly understate the price-responsiveness that technology and education will evoke in future years.

### *3. Augmenting TOU with Demand Charges*

Because TOU rates don't capture the wholesale price variation within a time block, TOU pricing is often combined with a separate charge for peak usage. These "demand charges" are a price per kilowatt for based on the customer's maximum usage (typically, during a 15 minute interval) during the billing period (usually a month), regardless of whether that usage occurs at a time when the system as a whole has a tight supply/demand balance or not. Most of the meters that register maximum usage for demand charge billing are not capable of storing information indicating the precise date and time at which that maximum usage occurred.<sup>9</sup>

Demand charges are a way to charge for a customer's peak usage, but the economic incentives that they establish are a very imperfect proxy for the real economic cost imposed on the system. First, demand charges are not synchronized to the usage on the system as a whole, so they charge as much for a peak usage that occurs at a lower demand time as at a higher demand time. In fact, this might not be as bad as it seems at first, because heating or air conditioning-driven demand peaks tend to be highly correlated across users within a region. More importantly, however, by charging only for the peak usage, demand charges don't give strong (or potentially any) incentive for a customer to conserve until usage is near the peak level for the period. If a very hot day occurs early in a billing period, the demand charge may give a customer little incentive to conserve after that.

The economics of demand charges made more sense under traditional utility regulation. The concept was to charge customers for their contribution to the need to build additional peaking capacity (though in practice this still suffered from the problem that the customer's peak consumption may not coincide with the overall peak consumption).

Apart from the peak, charges varied little. This makes much less sense in a deregulated wholesale market where demand increases result in significant increases in wholesale price even before the system gets right up to its capacity.

In addition, demand charges make no adjustment for the supply side of the market. If an unusually high number of forced outages occurs on a moderately hot day, the system can be more strained than on a very hot day, even if the total system load is lower. Though wholesale prices vary systematically with system demand, many other factors cause wholesale prices to fluctuate throughout the month. Variations in supply availability (and the prices at which that supply is offered) can be as important as variations in demand in explaining fluctuating wholesale prices. Thus, although demand charges do enhance the ability of TOU pricing to reflect true economic costs of service, they still fall well short of RTP. The demand charge approach grew out of the technological metering limitations that existed many decades ago and is now obsolete.

#### *4. Interruptible Demand Programs*

The physical attributes of electricity systems imply that excess demand cannot be rationed using the standard non-price mechanism: queuing. Instead, a systemwide excess demand can lead to a collapse of the entire grid, cutting off supply to all users. Thus, if for some reason economic incentives fail to equilibrate supply and demand for even a brief period, the system operator must have the ability to curtail usage by some customers. The response has been interruptible demand programs, which give the system operator the right to instruct the customer to cease consumption on very short notice. In return, the customer usually receives a reduction in its flat (or TOU) electricity rate, or it receives a periodic fixed payment.

Along the spectrum of time-varying price plans, interruptible electricity rates are nearly at the opposite end from RTP. These rates are constant nearly all of the time. When the system operator declares certain potential shortages, however, these customers are called upon to cease electricity consumption entirely. Despite the name, service to these customers is not actually physically interrupted in most cases. Rather, the price that they face increases dramatically. In one program in California, customers on interruptible rates

were required during declared shortages either to stop consuming or to pay \$9.00 per kWh for their continued consumption, a more than 40-fold increase. Thus, although these programs appear at first to be a departure from using a price system to allocate scarce electricity, in fact most programs are simply a crude form of RTP.

Seen in this light, interruptible programs are RTP with very blunt price changes: very poor granularity or timeliness unless a system emergency occurs, in which case the rate increases – for short periods and on short notice – so much that nearly all contracted customers choose to stop, or drastically reduce, consumption.<sup>10</sup> Interruptible programs offer a certain amount of insurance to customers, because they are told that they won't be called more than a pre-specified number of times during a year. As I discuss below, however, price protection products can supplement RTP to offer at least as much insurance.

### *5. Critical Peak Pricing Programs*

A recent innovation in time-varying pricing is critical peak pricing (CPP), which has some attributes of RTP and some of interruptible programs. CPP programs usually start with a TOU rate structure, but then they add one more rate that applies to “critical” peak hours, which the utility can call on short notice.<sup>11</sup> While the TOU program has poor granularity and timeliness, as discussed above, CPP allows a very high price to be called on very short notice, thereby improving both aspects of the rate structure. Thus, CPP is similar to interruptible programs except prices are not set so high as to cause most customers to reduce consumption to zero. In practice, interruptible programs are usually offered only to large customers, while CPP is envisioned to be used much more broadly.

CPP programs typically limit the utility to call no more than 50 or 100 critical peak hours per year. CPP is a clear improvement on TOU with demand charges, because the additional charges are based on consumption when the system is actually constrained, rather than when the particular customer's demand peaks. CPP has some of the advantages of RTP, because retail prices are allowed to vary with the wholesale market. Of course, CPP is much more constrained than RTP: the CPP peak price is set in advance and the number of hours in which it can apply is limited. A modification of the single-peak CPP is CPP with two callable peak retail prices, such as 50¢ or \$1.00 per kWh, so that the utility has

more flexibility in the strength of the retail price incentives it can utilize.

CPP programs are the natural evolution of demand charges when more sophisticated metering is available. Charges increase at critical system peaks rather than at the individual customer's demand peak, which is much more consistent with the true costs of consumption. CPP still has two economic weaknesses, though they may actually be strengths in terms of customer acceptance. First, the prices are limited and levels are preset for the critical peak periods, therefore they can't be calibrated to move with the actual prices in the wholesale market. Second, the number of critical peak hours that can be called in a year is limited. As a result, the utility protects customers against seeing very high prices, even only on marginal purchases, for more than a fixed number of hours. As discussed below, RTP can be designed to offer much the same level of hedging, while still giving the customers strong incentives on the margin to conserve when the market is tight.

#### *6. Real-time Demand-Reduction Programs*

While all of the pricing approaches discussed thus far charge customers time-varying prices, demand-reduction programs (DRPs) pay a customer to reduce consumption at certain times. A customer signed up for a DRP is eligible to be contacted by the utility or system operator with an offer of payment in return for the customer reducing consumption.<sup>12</sup>

These programs must first determine a baseline from which demand reduction can be measured. Once the baseline is set, the price offered for demand reduction determines the level of economic incentive to reduce demand when the system operator calls.

Much like CPP, and unlike TOU, real-time demand-reduction programs attempt to recognize the idiosyncratic daily and hourly variation in system stress and give customers incentives to respond. Demand-reduction programs are activated by the system operator when grid conditions meet certain pre-determined criteria that indicate that the supply/demand balance is likely to be very tight over some ensuing period of time. The operator then offers to pay participating customers to cut back their usage. In general, these programs are fairly blunt instruments; the system operator simply announces when the program is in effect. The price offered is usually pre-determined and does not vary with the tightness of supply. In this way, DRPs might be thought of as just the mirror



image of critical peak pricing and, like CPP, an improvement in both price granularity and timeliness compared to TOU pricing. Unfortunately, DRPs have a significant flaw that make them much less effective than CPP.

The fundamental weakness with demand-reduction programs is that there is no reliable baseline from which to pay for reduction.<sup>13</sup> With most goods, the natural baseline is zero: you start with none of the good and pay more as you consume more. Programs that pay for demand reduction generally set a baseline that comes from the past behavior of the customer. The baseline-setting process creates two serious problems.

First, if the program is voluntary, it will be joined disproportionately by the customers that already know they will have lower consumption relative to their assigned baseline. For instance, if the program uses last year's consumption as the baseline (perhaps with an adjustment for weather), the companies that have shrunk since last year will be the first to sign up. Their electricity consumption has fallen compared to the baseline for reasons having nothing to do with the program. The operator ends up paying for "conservation" that would have occurred anyway. Meanwhile, the companies that have grown rapidly since last year simply won't sign up.<sup>14</sup>

Second, if the baseline that is used can be affected by the customer, it will probably discourage conservation during times when the payments are not in effect. For instance, consider a plan that sets the baseline at the level of consumption the customer had on the previous day. Then on days when the payments are not in effect, customers would be foolish to conserve at all since that would just lower their baseline. Californians saw this effect in the 1970s with water rationing. Many users figured out that they were better off being profligate in normal-rainfall years so that they would have a higher baseline if a drought hit.

To overcome these problems, there could be a program that uses a baseline from an earlier period and is not voluntary. However, this would raise serious equity concerns. Shrinking companies would reap a windfall and expanding successful companies would be penalized.

Ultimately, real-time demand-reduction programs are very imperfect substitutes for critical peak pricing or RTP. They require the same metering technology as critical peak

pricing and approximately the same level of information transmission from the system operator. Demand-reduction programs limit the customer's liability by starting from a flat rate and reducing bills from there, but such bill-protection can be done easily within a CPP or RTP program, as described in the next section. Unlike CPP or RTP, demand-reduction programs suffer significant problems and potential conflicts in setting baselines.<sup>15</sup>

Lastly, although demand-reduction programs are often favored as a positive reinforcement mechanism, all the money that is paid out in positive reinforcement has to come from somewhere. Paying for demand reduction is not a free lunch. It most likely comes from higher general rates than would be necessary to reach the same revenue requirement under CPP or RTP.<sup>16</sup>

### **III. Issues in Implementing Real-time Pricing**

I will now focus on real-time pricing in discussing four important issues in implementation: risk to customers and retailers, distributional impacts of adopting RTP, mandatory versus voluntary implementation, and the role of demand response in setting and meeting reserve requirements. TOU pricing already exists and is used widely, but does not give economic incentives that are particularly accurate, as discussed above. Much of the discussion that follows applies to other time-varying retail price programs, including critical peak pricing and pay for demand reduction. As discussed in the prior section, interruptible programs are extreme and extremely blunt versions of RTP; their usage would be minimized in a well-functioning electricity system.

#### *1. Mitigating Customer Price Risk and Retailer Revenue Risk*

For more than 75 years, customers of regulated electricity utilities were protected from price volatility by a regulatory system that permitted rates to change only very infrequently. Regulated utilities met their revenue requirements over the long run by adjusting rates monthly or annually, with regulatory approval. Because utilities produced most of the power that they delivered, the primary risk they faced was from fuel price changes. Even that risk was reduced over the last 30 years, actually shifted to customers,

with the increasing use of automatic fuel adjustment clauses.

Now that more utilities (and other retail electricity providers) are net buyers of power, much of the utility-retailer cost risk comes from volatility in the wholesale price of power. Under a simple RTP system the retailer transfers that risk to customers by setting a retail RTP price based on the wholesale market price. If the retailer owns generation of its own or has other costs unrelated to wholesale power, however, it will still face risks in meeting its own revenue requirements. Thus, both potential customers and utility-retailers have expressed concern about the risks that may accompany RTP.

There are a number of different models for mitigating customer price risk and retailer revenue risk under time-varying retail prices. They differ in two basic dimensions: (1) customers can actively participate in hedging of price risk or the retailer can carry out hedging on behalf of passive customers, and (2) other costs unrelated to power consumption (such as stranded investments, and transmission and distribution costs) can be incorporated through adjustments to time-varying retail power prices or through lump-sum mechanisms. Capital gains and losses from retailer hedging are among the costs (or gains) to be incorporated.

The historical flat retail electricity pricing represents the extremes in both of these dimensions. All retail price hedging is carried out by the utility, which offers only the very-hedged flat-price product, and customers are completely passive. The utility meets its revenue requirement completely through adjustment of the flat rate, which covers both fixed and variable costs. To the extent that the retailer hedges its wholesale (or fuel) price risk, the capital gains or losses associated with those sunk gains or losses are incorporated when the flat rate is changed. TOU programs likewise limit the customers' retail price risk by drastically reducing the volatility of retail prices, and meet revenue targets by adjusting these rates gradually over time. Similarly, demand reduction programs limit customer risk by guaranteeing that the customer never faces a price above the posted tariff, and underwrites the program with higher flat or TOU retail rates.

#### *RTP with a Customer Baseline Load (CBL)*

A number of RTP programs, including the largest U.S. program, operated by Georgia Power, limit customer price risk by assigning each customer on RTP a baseline consumption

level that the customer is required to purchase at the regulated rate during each hour. The price the customer pays for its baseline consumption is usually the regulated TOU rate that the customer would otherwise face. Starting at the baseline, the customer then pays the real-time price for any consumption above its baseline level and receives a rebate based on the real-time price if its consumption falls below its baseline level. In financial terms, the baseline is just a forward contract for a quantity equal to the customer's baseline level, *i.e.*, a hedge contract, which the customer has purchased at a price set by the regulatory process. Such programs are generally called two-part RTP programs with a customer baseline load (CBL).<sup>17</sup> Under RTP with a CBL, customers are passive in the hedging activity, though Georgia Power offers additional hedging instruments that the customer can actively choose to purchase.

The CBL approach also creates a mechanism for covering costs that are not directly related to the incremental cost of energy without distorting the RTP price signal. By setting the CBL and the rate at which that energy is purchased, the utility (or regulatory agency) builds in a fixed charge for a fixed quantity of energy. That fixed charge can be used to cover costs (or give refunds) without distorting the incremental energy price. For instance, if the regulator decided that the utility was permitted to raise an additional \$1 million to cover a stranded investment, an increase to the TOU rate would increase revenue earned from customers on RTP with a CBL by raising the cost of the baseline power. Importantly, it would not affect the price that the firm pays for power on the margin.

The California crisis, however, made clear a significant difficulty with the CBL approach. If the customer buys the baseline at a price that is not closely related to the expected market price, purchase of the baseline quantity amounts to either a subsidy or a tax depending on whether the regulated price is above or below the expected market price. That alone would create an equity issue, but in practice it also creates a significant influence and lobbying problem.

An RTP plan with a CBL was proposed in California during the spring of 2001. At that time, the real-time price was expected to be well above the regulated price at which customers would purchase their CBL. Once companies understood this (the plan was to be only for large customers), their focus turned almost entirely to lobbying for a high CBL

for themselves.<sup>18</sup> Any RTP program with a CBL will include an implicit transfer payment to or from the customer so long as the regulated price differs from the expected real-time price. Thus, baselines set by any regulatory process will be subject to intense lobbying and related influence activities.

*RTP with Build-Your-Own Baseline*

There is a device similar to the CBL that can avoid the lobbying and influence problems: the forward contract that the CBL itself was meant to mimic. One way to think of this is as real-time pricing with a build-your-own (BYO) baseline. As the name suggests, customers take an active role in determining the extent to which they hedge price risk. Rather than being assigned a CBL for each hour of the year, the customer can purchase a baseline quantity, *i.e.*, a forward contract, for each hour in order to hedge as much price risk as it wants. The key would be for the retailer to offer the baseline or forward contract for each hour at a price that equals the best forecast of that hour's future spot price. Georgia Power, one of the first users of the CBL, now offers such futures contracts for RTP customers that want more hedging than is provided by the CBL. They refer to these as a price protection products. The advantage of this approach is that because the BYO baseline is purchased at a price that reflects the expected real-time price, unlike the CBL, it contains no subsidy or tax on average. It is simply a risk-hedging device. This approach also makes it easy for the retailer to offset the baseline it sells forward by purchasing power forward in the wholesale market at approximately the same price at which it sell power forward for the BYO baseline.

The BYO baseline allows customer to avoid purchasing most of their power at the real-time price. Yet, for incremental consumption decisions, the customer still faces the real-time price as its cost (or opportunity cost, since it can resell power it doesn't use from its BYO baseline), and thus has strong incentives to conserve at peak times. The BYO baseline cannot create a perfect hedge since the customer won't know in advance exactly what quantity it will want to consume in each hour. To the extent that customers want even more stable electric bills than the BYO baseline would provide, other products, including call options, may develop in a private market. In fact, the BYO baseline itself could be provided by a private market. In reality, however, regulatory uncertainty at

the beginning of an RTP program may make private parties reluctant to enter this retail hedging market, so it would probably make sense for the utility to be the first provider of such risk-hedging products.

One disadvantage of the BYO baseline in comparison to the CBL is that the BYO baseline cannot be used to cover sunk or fixed costs. The forward price offered must be very close to the expected real-time price when customers voluntarily choose how much baseline to purchase. The BYO baseline would have to be augmented with some other fee (or rebate) mechanism in order to assure that revenue requirements are met. The CBL implicitly creates such a fee mechanism by assigning each customer in the program a baseline level of consumption that is required to be purchased at a regulated price that is not tied to the expected spot price. An alternative is to use a fixed adder on the RTP price to cover sunk or fixed costs, as I discuss in the next subsection.

#### *RTP with Retailer Hedging on Behalf of Customers*

Even if customers are on a simple RTP plan without any baseline and retailers are buying most of their power in the wholesale market, retailers can still stabilize RTP customers' bills by hedging on their behalf and using the profit or loss from such hedging to offset power price fluctuations. In this approach customers are completely passive, putting all hedging decisions in the hands of the retailer. By purchasing some power on long-term contracts, the retailer hedges wholesale power costs. The question this creates is how the retailer can pass through the profit or loss from hedging in a way that minimizes price distortions.

Historically, utilities have hedged on behalf of their customers by generating most of their own power, but they have passed through the results of this hedging through occasional adjustments to flat retail rates, a process that creates distorted incentives to customers. In restructured markets, the practice has continued: by retaining ownership of generation, signing long-term power contracts, or buying generation inputs forward, retailers still hedge power prices. And they've still passed along the outcome of their hedging through occasional adjustment to a flat retail rate. Retailers, however, can hedge on behalf of RTP customers and can pass along the outcome from hedging without undermining the volatility of RTP prices by attaching a constant (over the billing period) adder or

subtractor to the RTP price.

To see this, begin by assuming that the retailer engages in no hedging. It charges customers a fixed per-kWh transmission and distribution (T&D) charge plus the spot price of energy in each hour.<sup>19</sup> In this case, the customers' monthly bills, and retailer's revenues, would be as variable as the month-to-month variation in the weighted-average spot energy prices.

To attain the goal of monthly bill stability, the retailer would sign a long-term contract to buy some amount of power at a fixed price.<sup>20</sup> To fix ideas and keep the presentation simple, assume that the retailer signs a long-term contract at the same price for each hour. Such a contract will be at about the average spot price of the electricity that the parties anticipate over the life of the contract, but in any given month the contract price could be greater or less than the average spot price. If there are significant shifts in the market, the contract prices could end up being much cheaper or more expensive than the spot wholesale prices.

This contract can be considered a financial investment that is completely independent of the retailers service to final customers. The critical point is that the retailer's return on this financial investment varies directly with the average spot price of energy, and that return can be applied to change the average level of customer bills. When viewed this way, it becomes clear that the long-term contract can be used to affect the average price level without dampening the price variation. The gains (when the average spot price is higher than the contract price) or losses (when the average spot price is lower than the contract price) from the long-term contract could be distributed to customers with a constant surcharge or discount on each kWh sold during that billing period.

The retailer would then charge the customer the spot price plus or minus an adjustment equal to the average return per-kWh from the long-term contract. Since the profit earned from holding the contract is greater when the spot price is higher, it would be used to offset the high *average* spot price, thus lowering the volatility of monthly electricity bills. If the retailer hedged nearly all of the load it served, then this offset would be sufficient to nearly eliminate variability in monthly bills. If the retailer hedged, say, 80% of the load, then about 80% of the variability in monthly bills would be eliminated. The actual price

that the customer was charged in each hour, however, would still have the same volatility as the spot price. The critical point is that the gains or losses in any given hour from the forward contract need not and should not be collected in that particular hour. There is no economic argument for doing so and doing so would greatly reduce the variability of retail prices and, thus, the economic incentive for conservation.

Figure 5 illustrates the effect that this approach would have had in California during June of 2000 if the retailer had been lucky enough to sign a long-term contract before prices increased. In this illustration, the retailer is assumed to have signed a contract for 80% of its load at 6¢ per kWh. In addition to energy charges, the retailer is assumed to assess a 4¢ per kWh charge for transmission and distribution. The T&D charge is added to all prices for ease of comparison. The three horizontal lines show the load-weighted average price a customer would pay (assuming it had the same load profile as the system as a whole) if the retailer were fully hedged (lowest line), if it were completely unhedged (highest line), and if it were 80% hedged (middle line).

[FIGURE 5 ABOUT HERE]

Of the two volatile lines, the higher shows the real-time price a customer would pay with no hedging and the lower shows the price the customer would have paid if the retailer had combined real-time pricing with 80% hedging, in this case, purchased through long-term contracts. The load-weighted average of the higher line is 18.08¢, the same as the highest horizontal line. The load-weighted average of the lower line is 11.62¢, the same as the middle horizontal line.<sup>21</sup> This illustration demonstrates that under real-time pricing with long-term contracts, a customer could face the same volatility in prices as it would under RTP with no hedging. The only difference would be that the price curve would be shifted down by the “profits” from the long-term contract, which in this example are 6.46¢ per kWh. During the hours of extremely high spot prices, customers would face prices nearly as extreme, and would have a strong incentive to reduce consumption. Yet, the average monthly prices (and monthly bills) the customer would face would be much less volatile than without hedging.<sup>22</sup>

Hedging by the retailer is particularly valuable in a regulated context where customers do not have the choice to hedge price risk themselves by signing private contracts with



generators or other market participants. In a competitive retail market, retailers would probably not arbitrarily impose a level of hedging and retail price. Instead, they would offer to pass through the wholesale spot electricity price and would augment that offering with various price protection programs, such as a BYO baseline. The retailer would then hedge its wholesale price risk in a way to match the retail price hedging that its customers have chosen to purchase. In essence, the retailer would serve as a broker of risk hedging services.

More generally, a fixed per-kWh adder or subtractor can always be used to attain any desired average retail price while maintaining the time-variation in retail prices.<sup>23</sup> The issue of time-variation in retail pricing can be completely decoupled from the issue of targeting a given average price or, a given revenue requirement. Thus, the same approach can be used with BYO baseline to recover costs unrelated to wholesale power prices, including payments for stranded investments and transmission and distribution.

This is, of course, not the ideal way to change average bills, since it means that retail prices will not exactly reflect the wholesale market costs, creating deadweight loss. The economic ideal would be to pass through wholesale prices exactly in each hour and separately make lump-sum adjustments, unrelated to consumption during the month, in order to hit an average price target. The problem is that lump-sum adjustments, as illustrated by the CBL, raise significant equity issues and lobbying costs. Furthermore, the size of the deadweight loss from mispricing increases in proportion to the square of the mispricing. The adder or subtractor envisioned here would probably be quite small compared to the variation in wholesale prices. Thus, while it would create small mispricing in many hours, it would allow a system that prevents the large mispricing that occurs during peak hours absent RTP.

### *3. Distributional Impacts of Adopting RTP*

Of great concern in discussions of RTP programs is which participants will be winners and which will be losers from adoption of such a change. A significant effect of RTP in the short run is to reduce the total payments to generators in the wholesale electricity market. This occurs because peak demands, when prices are very high (even if the wholesale market

is competitive), are reduced. In the long run, this means fewer new power plants need to be built and their costs need not be borne by customers. It is important to note these effects can benefit all customers, those on RTP and those who are not. Furthermore, if the wholesale market is not completely competitive, RTP reduces the ability of sellers to exercise market power, which further reduces wholesale prices and, again, benefits all customers whether on RTP or not. By creating real-time demand elasticity, RTP also lessens the likelihood of system shortages and the potential for customer interruptions or rolling blackouts.

Among customers on RTP, the benefits vary depending on when a customer consumes power. Customers with a relatively flat consumption profile – whose consumption has less hour-to-hour and day-to-day variance than others on RTP – will see the greatest benefits. Because their consumption does not “peak” as much as the aggregate demand at high-demand times, they will be buying a smaller share of their power at the most expensive times. As a result, their average price per kWh will be lower than for customers with more “peaky” demand.

Customers with more “peaky” demand, who do not respond to the higher retail prices that will occur when supply is tight, will consume a disproportionate share of their power at the more expensive times. These customers will face a higher average price per kWh than will customers with a flatter demand profile. In California, the peak demands occur on hot summer days and are driven primarily by air conditioning. There are no completely reliable rules, but customers that use power disproportionately for air conditioning (or for heating in areas that peak in the winter) generally have more “peaky” demand that is greatest at the time when electricity is more expensive. Customers that use power disproportionately for other uses (lighting, industrial production, machinery) generally have flatter demand profiles.

Customers with “peaky” demands might end up paying more than under flat rates, but it is quite possible that they would not. While they would pay high prices at peak times, the overall efficiency of the system would improve and there would be less idle capacity that all customers must pay for in the long run. The average price would decline with RTP, but these customers would consume a disproportionate share of their power

at higher-price times, so it is hard to know for sure whether their bills would increase or decrease. Certainly, to the extent that they can respond to high prices by conserving or shifting demand, their bills would be more likely to decrease.

Many industrial customers that run 24 hours a day, 7 days a week have expressed concern about RTP because they say that they cannot reduce their demand or shift usage. Customers that run 24x7, however, tend to have flatter demand profiles than the system as a whole. Thus, although they would have to buy power at higher prices during system peaks under RTP, they would buy a smaller share of their power at these times than would other users. Their overall power bills would decline.

The reason for this is that the system is not driven by one customer's ability to conserve, but by the ability of users in aggregate to lower demand at peak times, thereby reducing wholesale electricity prices and reducing the risk of rolling blackouts. If a certain customer cannot conserve at all without suffering major economic losses, then RTP offers enormous benefits, because it reduces the chance of rolling blackouts, which are especially costly to such customers, and it reduces the wholesale price of electricity at peak time because other users reduce their demand.

In nearly all areas where RTP is being considered or has been implemented, it has not been offered to small commercial or residential customers. It certainly makes sense to start with large customers, because the cost of metering them is trivial compared to the potential efficiency gains. But it should be recognized that customers who remain on flat-rate service are imposing demands at peak times that they are not paying for. In the following section, I discuss a strategy for ensuring that RTP customers are not saddled with power costs attributable to flat-rate customers.

On the other hand, some industrial customers have suggested that if they switch to RTP they should be rewarded since their behavior benefits other consumers at peak times. While that is true in the short run, it is only at the expense of producers. The price reduction at peak times is a transfer from producers to consumers, not a net benefit to society. The overall saving to society from a RTP customer reducing its consumption by one kWh is no more than the price of that kWh.<sup>24</sup> If prices are set appropriately for both RTP and flat-rate customers, as I describe in the next section, then being an RTP

customer is its own reward and should not require further subsidy. The proper policy is to end implicit subsidies in flat-rate retail pricing, not to create even larger offsetting subsidies for customers on RTP.<sup>25</sup>

#### *4. Mandatory Versus Voluntary Implementation*

A question that arises frequently in discussions of RTP is whether it will be mandatory or voluntary for a given class of customers. Since virtually all consideration of RTP has been only for large customers, the class of customers for whom RTP might be mandatory is those large customers. In practice, all of the programs in the U.S. are voluntary.

It is important to remember that a mandatory RTP program does not mean that customers need to be exposed to the full risk of the spot electricity market, since hedging by the retailer or by the customer can greatly reduce the price risk associated with RTP. A more accurate description would be a “default” RTP program: customers by default pay the real-time price, but they can sign forward contracts that mitigate the volatility in costs they would face if they purchased all of their power at the spot price.

Still, many market participants argue that if any RTP program is instituted it should be done on a voluntary basis. While a voluntary program is attractive in many ways, it also raises difficult equity and efficiency issues that don’t arise if RTP is the default.

The customers that would sign up for a voluntary RTP program would be those that have a more attractive load profile – either flatter than most or actually peaking at times when the system demand is low – or those that *would have* a more attractive profile once they responded to real-time prices. These customers would get lower bills if they paid the wholesale price of the power they used under RTP rather than the average wholesale price of the power that all customers use. In large part, a voluntary program would identify the customers who have been subsidizing other users that consume more at peak times.

This, however, creates a problem for the retailer. In order to avoid a revenue shortfall, either it must charge a higher average price to those that don’t sign up for RTP than to those that do or it must place an adder on the RTP price so that the RTP group as a whole faces the same average price as non-participants.<sup>26</sup> A lower average price for RTP participants would be consistent with the cost of serving this group, but it would

very likely raise difficult political issues. Rather than being seen as an end to a historical cross-subsidy, it would likely be portrayed as a new subsidy to RTP customers.

On the other hand, an adder to RTP prices that equalizes the average RTP price and the average non-RTP price would undermine the incentives to join the program. If the average per-kWh price for RTP customers is the same as for non-RTP customers, then some – probably about half – of the customers on RTP are paying a higher average per-kWh price on RTP than they would if they switched to the non-RTP rate. Those customers would be better off switching to the flat rate. As they did, this would reduce the average per-kWh rate among those remaining on RTP. In order to equalize prices between RTP and non-RTP customers, the retailer would have to raise the RTP adder. Once it did so, a new set of RTP customers would find they were paying a higher average price than they would on a flat rate. They would then be better off switching to the flat rate. This RTP death spiral would continue until only the customer with the most attractive load profile is left on RTP and the adder is set so that this customer is just indifferent between staying on RTP or switching to a flat rate.

An alternative to the RTP death spiral would be to institute RTP on a voluntary basis, but charge each group, RTP and flat-rate, an average price equal to the average cost of serving that group.<sup>27</sup> That is, this would be a no-cross-subsidy implementation of a voluntary RTP program.

When first offered, the RTP plan would primarily attract customers with the most attractive load profiles. But once these customers switched to the RTP plan, the average cost of serving customers that remained in the flat-rate plan would rise. This would make the flat-rate plan less attractive relative to RTP and would cause more customers to switch. Those that would now want to switch would be a selection of customers remaining in the flat-rate plan that have relatively more attractive load profiles. Again, the switching would cause the average cost of serving flat-rate customers to rise, and thus would increase the price charged to them, making the flat-rate plan even less attractive. While this approach might not cause a complete unraveling of the flat rate, it would almost surely attract significant business and result in a higher average price for those that remain on flat rate than for those that move to RTP.

Again, it could still lower the price charged to both groups compared to no RTP since the incentive to consume less at peak times would reduce overall peak demand and wholesale prices. It is worth pointing out that the no-cross-subsidy implementation of a voluntary RTP plan might also be achieved without a regulatory process. Retail choice is likely to have this effect if it is implemented in a way that does not offer a subsidized default utility rate. Customers that are cheaper to serve will abandon the flat-rate utility service in order to get the lower prices associated with their more attractive load profiles. As they do, if the remaining utility customers are not subsidized, their flat-rate price will rise above the average price paid by customers who have moved to a competitive retailer offering RTP.

Implementation of retail choice thus far, however, has generally offered customers remaining with the utility a fixed price independent of the costs of serving these customers. This has set up two types of incentives. First, it has reduced the attraction of moving to a competitive retailer, because as some customers have moved, the price to the remaining customers has not been affected, in contrast to the no-cross-subsidy implementation I've just described.<sup>28</sup> Second, it has created incentives for customers to jump back and forth between competitive suppliers and the fixed-price utility offering. When wholesale prices in California jumped in summer 2000, competitive retailers raised their prices to reflect that increase. The result was that many customers returned to the fixed-price utility rates, which were then being heavily subsidized. When wholesale prices fell in summer 2001, many of the same customers jumped back to competitive retailers.

Successful implementation of RTP does not depend on making it compulsory for any group. It can be implemented on a voluntary basis if it is done with a commitment of no cross-subsidy between RTP and flat-rate customers. In fact, such an approach would be likely to create its own momentum as low-cost customers abandon the flat-rate and increase the wedge between the average prices paid by RTP and flat-rate customers. This would be the result of retail competition if prices to customers that remained on the utility's flat-rate plan were reset to reflect the cost of procuring power for those customers. In practice, however, a voluntary program is usually implemented with price, or subsidy, protection for customers that choose not to switch. This dampens the economic incentive to switch to

RTP. The problem is greatly exacerbated if the flat-rate customers are also given a fixed rate that doesn't change when average prices in the wholesale market changes, as occurred in California during the summer of 2000.

##### *5. Using Demand to Meet Reserve Requirements: Economics meets Engineering*

Administrators of nearly all restructured electricity markets, and of many that are still fully regulated, have suggested that demand response to time-varying retail prices could be used to help meet reserve requirements. The Federal Energy Regulatory Commission (FERC) has endorsed this concept in its draft Standard Market Design. Implementation of this approach, however, exposes a conflict between the economic analysis of electricity markets and the operating procedures that electrical engineers use in operating the grid. The grid operator wants to have resources that it knows with near-certainty it can call on to increase supply or reduce demand.<sup>29</sup> Grid operators tend to be resistant to the idea of equilibrating supply and demand by increasing the price in a spot electricity market and hoping that demand in aggregate will respond.

The grid operators concerns are understandable since any supply/demand imbalance that lasts even a split second can be catastrophic. They are less attracted to the idea of balancing supply and demand through price adjustments that yield small quantity changes from thousands of customers, because none of those customers pre-commits to make a specific change under specific conditions. Rather, the responses to price changes are probabilistic, and the reliability of the aggregate response is due only to the law of large numbers applied to many independent buyers. To be concrete, if demand is exceeding supply and adjustment is supposed to occur through a price mechanism, a grid operator has no one to call to assure that demand response occurs.

It is for this reason that grid operators tend to be much more supportive of interruptible demand and contracts for pre-specified demand reduction than they are of critical peak or real-time pricing as a method for balancing supply and demand. Grid operators, however, can be satisfied while still allowing price response to play a major role supply/demand balancing. Interruptible and demand-reduction contracts will still have a role to play as a backup mechanism. Given the engineering realities of electricity systems, the system

operator has to have some ability to cut off customers when the system is nearly overloaded and normal market processes either have failed or cannot respond quickly enough. It makes much more economic sense to establish priorities for such situations in advance and compensate those who volunteer for curtailment, rather than choose customers or areas randomly, as was done in California during early 2001.

Still, it is important to recognize that reducing demand through customer interruption or large pre-specified demand reductions is likely to take a much larger economic toll than is time-varying retail prices: forcing 100 customers to reduce consumption by 100% will be much more costly than inducing 10,000 customers to each reduce consumption by 1%, on average. Grid operators have long recognized that it isn't practical to enter into thousands of contracts for small demand reductions, or to call each customer to ask for reduction, but the price mechanism is the tool that can achieve that outcome without formal contracting. If market processes are allowed to operate effectively, interruptible programs and demand-reduction contracts will be an absolute last-ditch mechanism that will hardly ever be used.

Currently, reserve requirements are established based on forecasts of system peak demands that do not take into account demands response to price changes. Implicitly, demand is assumed to be completely price-inelastic and system reserve requirements are set to be some percentage, usually between 10% and 20%, above the forecast system peak. Contracts for demand interruption or reduction fit easily into that paradigm. Incorporating the price-responsiveness of demand under RTP or CPP requires a change in the paradigm. It overstates the effect of RTP to say that no reserves are needed once RTP is implemented, but it understates the impact to say that RTP would simply reduce the system peak so the same reserve requirement should be applied to a somewhat lower peak forecast. Time-varying retail prices demand will mean not only that the system peak will be lower, but also that the necessary percentage reserve level will be reduced, because an unexpected system shortage will be addressable in part through the response to price increases. When it is first implemented in a system, the reliability of price response will be unproven, so the operator will rely very little on it to meet reserve requirements. Over time, however, the response will be more reliably forecast and price-responsiveness will be able to take on an increasing role in assuring system reliability. Though it will never fully replace other



forms of reserves, price-responsive demand will eventually cut the necessary reserve levels substantially.

#### **IV. The Long-Run Impact of Real-Time Pricing**

It is clear from the previous section that conversion to real-time pricing of electricity would require changes both in the technology of metering and billing and in the financial transactions associated with electricity consumption. It is important to examine the benefits of RTP to determine whether they justify the costs. To do so, I have simulated a competitive electricity market in the long-run with plausible demand elasticities, load shapes, and production technologies, assuming free entry and exit of generation.<sup>30</sup> These simulations are not intended to be a precise reflection of any particular market, but are intended to illustrate with plausible parameters the magnitude of impacts that RTP would have.

The first difficulty in such simulation of the impact of RTP is that it is dependent on the demand response that the price fluctuation will elicit. Estimation of electricity demand elasticity has a long record, but the experiments that have been used as the basis for estimation have significantly constrained inference about how an RTP program would change behavior.

Most of the estimates that exist are based on pilot programs that were run by utilities in the 1970s and 1980s. Two factors make these programs and the resulting econometric studies inapplicable to forecasts of the impact of RTP. First, most of the programs were pilot studies of residential TOU pricing, not RTP. Those that were RTP-like programs mostly used blunt price changes (more like critical peak pricing). Second, and more important, the response to RTP is largely dependent on the technology available for automated response. Few customers would find it worthwhile to manually adjust most electricity usage in response to real-time price fluctuations. The technology for automated response has improved drastically since the 1980s and continues to improve at a very rapid rate. Thus, to get a reasonable sense of the possibilities for demand response, we must look to recent experience, though even these probably understate the elasticity that will result as technology continues to improve.

The two most relevant programs for estimating elasticity of demand to real-time price variation are in the U.K. and Georgia. Patrick and Wolak (1997) studied response of industrial customers to real-time price variation in the U.K during 1991-1995. Braithwait and O'Sheasy (2002) studied the response of industrial customers on Georgia Power's RTP program that has been in existence since 1993. Unfortunately, even these recent studies give a wide range of elasticities, from -0.001 to -0.25 in Patrick and Wolak and from -0.01 to -0.28 in Braithwait and O'Sheasy. I use two figures in the simulations I present: -0.075 to represent price response in the shorter run, and -0.15 to represent price response in the longer run. The short-run response might be less elastic than -0.075 and the long-run response might be more elastic than -0.15, especially as the technology for automated response improves, but this will at least give an idea of the possible effects within a plausible range of elasticities.

The second important input to the simulation is a system load profile. I take the load profile of the California ISO system for the two-year period November 1998 to October 2000. While this is not perfectly representative, the first year of the period was moderately low demand and the second year was moderately high demand, so it does give a broad representation.<sup>31</sup>

The third type of inputs necessary to carry out such a simulation is the production technologies available. I assume that there are three technologies: a baseload technology (high fixed capacity cost, low variable cost) with approximately the cost properties of coal plants, a mid-merit technology (medium fixed capacity cost, medium variable cost) similar in costs to combined-cycle gas turbines, and a peaker technology (low fixed capacity cost, high variable cost) that reflects costs of a combustion turbine generating plant. Table 1 presents the exact cost assumptions used for each technology. I would not suggest that these are the best forecast of costs for these technologies, but they are in a plausible range. More importantly, the qualitative results are robust to a wide variety of assumptions about the exact costs, as discussed in Borenstein (2003).

[TABLE 1 ABOUT HERE]

To carry out the simulation, I first calculated the flat-rate tariff that would be required to cover the costs of the least-cost production of the assumed load duration curve, where

least-cost production was defined by the range of hours a plant is used per year, as shown in table 1. I assumed that the load duration curve taken from California would be the actual quantities demanded at this flat rate, \$79.13 per MWh (or 7.913¢/kWh). This yielded the top row in both the upper and lower panel of table 2. Note that because demand is inelastic in this case, capacity payments are required in order for firms to cover their costs. The flat-rate tariff is set at a level that produces enough revenue to cover those capacity payments, as well as operating costs.<sup>32</sup>

I then calculated competitive generation equilibria for each hour assuming that demand would be equal to its level in the assumed load profile if price were equal to \$79.13/MWh, but that it would change according to the assumed (constant) elasticity for higher or lower prices. Generators earn scarcity rents when the wholesale market clears above their marginal cost, and these scarcity rents provide efficient investment incentives for each production technology.<sup>33</sup> The lower rows in each panel of table 2 provide the resulting changes from the flat-rate base case in capacity of each generation type, costs, quantity consumed, and consumer surplus.

[TABLE 2 ABOUT HERE]

Table 2 reveals a number of effects of RTP. In the upper panel, it illustrates that RTP would likely raise the use of baseload technology slightly and greatly reduce the use of peaker technology.<sup>34</sup> Even a very slight elasticity of demand, -0.025, reduces the need for installed capacity by more than 10%. The shorter-run elasticity I posit of -0.075 would reduce the need for capacity by about 19% and the longer-run elasticity of -0.15 would reduce it by about 26%.

In this simulation, firms are assumed to be perfectly competitive, exactly covering their full costs (including a normal return on investment) in the long run. Thus, the only change in surplus in the long run is for consumers. The “Change in Consumer Surplus” column of table 2 shows the increase in annual consumer surplus and the adjoining column shows this change as a percentage of the customer annual energy bills prior to the introduction of RTP. Thus, for instance, introducing RTP in a market in which market demand exhibits an elasticity of -0.075 increases consumer surplus by about \$410 million dollars per year, or 4.42% of the pre-RTP energy bill.

In reality, RTP is unlikely to be introduced to an entire system simultaneously. Rather, it is likely that the largest customers would move to RTP first. In California, the state paid to have RTP meters installed for the largest 20,000 customers in the state, who together comprise nearly one-third of the state's demand. Assuming that the customers moving to RTP have about the same load profile as the system as a whole, moving one-third of customers to RTP can be approximated by examining a case with demand elasticity equal to one-third of the actual demand elasticity.<sup>35</sup> Examining the row with demand elasticity of -0.025, therefore, gives an idea of the effect of putting one-third of load on RTP when all demand has a price elasticity of -0.075.

Comparing these two rows, it is immediately apparent that putting one-third of load on RTP attains much more than half the benefit one would get from putting all of the load on RTP. In fact, the first third on RTP creates a consumer surplus increase of about \$234 million per year, the second third (from the -0.05 elasticity row) increases that figure by \$102 million per year, and the last third adds \$74 million in annual consumer surplus. This demonstrates the effect that many researchers have recognized, that there are diminishing returns to moving additional customers to RTP.<sup>36</sup>

At the same time, there are increasing costs of putting greater percentages of the load on RTP, because metering the largest users is cheapest per MWh. The costs of implementing RTP are difficult to pin down, because technology is constantly improving and because the costs may be greater or less than the direct metering costs. When California put RTP meters in the 20,000 largest customers representing nearly one-third of load, it cost about \$35 million. Some argue that the actual costs are greater than this, because utilities must create and install new software for billing and metering. Others argue that the utilities will actually save more than the costs of metering through reduced labor, because the new meters can be read remotely, without an onsite visit. In any case, recognizing that the RTP metering and other setup costs are one-time, or at least recur over decades, and the consumer surplus gains in table 2 are *annual*, it is clear that RTP for the largest customers easily passes a cost/benefit test, even when overall demand is thought to be fairly insensitive to price.

The results of these simulations are less conclusive on the net gain from metering the

remaining customers in a California-like system. The annual benefit from going from one-third to all load is estimated to be about \$176 million per year. Doing this would require installing about 10 million new meters in California. Some private companies have said they could install and operate the meters for a monthly charge of \$1-\$2 per meter. That would put the annual benefits in the same range as the annual costs. The other costs and benefits might then determine the outcome. How much would actually be saved on meter reading? How much additional cost would there be in billing operations, recognizing that systems would have already been put in place for having the largest customers on RTP.

The assumption of a -0.075 overall demand elasticity is probably reasonable in the shorter run, but is likely too low in the longer run, both because technology will improve over time and because customers will, with experience, become more savvy about how they can respond to higher and lower prices. The simulations assuming an overall demand elasticity of -0.15 may be more reasonable for the longer run, though I must emphasize that these figures are only intended to be in a plausible range, not precise estimates of how much demand response would actually occur.

With an overall elasticity of -0.15, putting one-third of the load on RTP can be approximated by the -0.05 elasticity row. The effects are larger than in the shorter-run case, but the additional impact from doubling the elasticity is only a 44% gain in the effect on consumer surplus from putting one-third of the load on RTP. The greater elasticity does mean greater gains from RTP, but the incremental gains from greater elasticity are declining.

Though these illustrations are useful in giving an idea of the potential gains from RTP, they don't take into account all aspects of electricity markets. First, there is no consideration of operating reserves. Including reserves would almost certainly move the analysis more strongly in support of RTP. RTP would decrease system peak loads, so using standard proportional reserve rules, it would reduce the amount of reserve capacity needed and the payments for that capacity. More importantly, RTP would increase the responsiveness of demand to system stress and thus would reduce the level of reserves needed for any given level of demand.

The simulations also have assumed that all generation is completely reliable, ignoring

outages. Outages would simply change generation costs if they occurred in equal proportions in all hours, but in reality they occur stochastically. Because significant quantities of capacity can randomly and simultaneously become unavailable, the variation in supply/demand balance is more volatile than would be inferred from looking at load variation alone. Greater volatility increases the value of RTP.

Lastly, the simulations also have ignored market power issues, instead assuming that free entry would bring a completely competitive market over the longer run. As Bushnell (2003) points out, price-responsive demand reduces sellers' ability to exercise market power. Reducing the exercise of market power not only reduces the large wealth transfers that it can engender, but also reduces the inefficient investment, allocation of production across plants, and consumption choices that result from market power. Including market power effects would almost certainly bolster the case for RTP.

## V. Conclusion

It is clear that time-varying retail electricity prices can significantly help to reduce system production or procurement costs, as well as to meet operating capacity reserve requirements. Many approaches to implementing this idea are in practice, some more widespread than others. RTP, the most attractive approach from an economic viewpoint, has faced two important impediments. The first, technological, has greatly diminished in the last decade.<sup>37</sup> The second, concern over price volatility, can be addressed effectively with hedging instruments or price protection plans, as has been demonstrated here.

Recognizing that these impediments have been overcome, RTP is clearly the right policy for medium to large customers. For small customers, including residential, metering costs and concerns about financial complexity might suggest a simpler approach. Some of the alternatives, particularly critical peak pricing, accomplish some of what RTP offers, while others, like Time-of-Use pricing, are a much smaller step towards improving the efficiency of electricity systems.

For more than 75 years, electricity planning has been based on meeting a peak demand that has been assumed to be price-insensitive. The payoff for greater demand-side price-responsiveness is reduced reliance on generation capacity that is operated only to

meet occasional demand peaks. The failure to incorporate demand flexibility has imposed unnecessary generating plants, production costs and environmental harm on society. An electricity system that permits adjustments on both the supply and demand side will improve efficiency, reduce costs, and benefit the environment.

## REFERENCES

- Borenstein, Severin. 2000. "Understanding Competitive Pricing and Market Power in Wholesale Electricity Markets," *Electricity Journal*, July.
- \_\_\_\_\_. 2001. "The Trouble with Electricity Markets and California's Electricity Restructuring Disaster," Program on Workable Energy Regulation Working Paper PWP-081, University of California, September. Available at <http://www.ucei.org/PDF/pwp081r1.pdf>.
- \_\_\_\_\_. 2003. "Estimating the Long-Run Impact of Real-time Pricing," Center for the Study of Energy Markets Working Paper, University of California Energy Institute, June.
- \_\_\_\_\_, Michael Jaske, and Arthur Rosenfeld. 2002. "Dynamic Pricing, Advanced Metering, and Demand Response in Electricity Markets." Study sponsored by The Hewlett Foundation and The Energy Foundation. Also released as Center for the Study of Energy Markets Working Paper #105, University of California Energy Institute, October. Available at <http://www.ef.org/documents/DynamicPricing.pdf>.
- \_\_\_\_\_ and Stephen P. Holland. 2002. "Investment Efficiency in Competitive Electricity Markets With and Without Time-Varying Retail Prices," Center for the Study of Energy Markets Working Paper #106, University of California Energy Institute, November. Available at <http://www.ucei.org/PDF/csemwp106.pdf>.
- Bushnell, James. 2003, "Looking for Trouble: Competition Policy in the U.S. Electricity Industry," this volume.
- Braithwait, Steven D. and Kelly Eakin. 2002. "The Role of Demand Response in Electric Power Market Design." Study by Lauritis R. Christensen prepared for Edison Electric Institute, October. Available at [http://www.eei.org/esg/policy/demand\\_response/demand\\_response\\_role.pdf](http://www.eei.org/esg/policy/demand_response/demand_response_role.pdf).



- Braithwait, Steven D. and Michael O'Sheasy. 2002. "RTP Customer Demand Response – Empirical Evidence on How Much Can You Expect," in Faruqui and Eakin eds. *Electricity Pricing in Transition*, Boston, MA: Kluwer Academic Publishers.
- Faruqui, Ahmad and Stephen S. George. 2002. "The Value of Dynamic Pricing in Mass Markets." *The Electricity Journal*, July, pp. 45-55.
- Faruqui, Ahmad, Joe Hughes and Melanie Mauldin. 2002. "Real-Time Pricing in California: R&D Issues and Needs," report prepared by the Electric Power Research Institute (EPRI) for the California Energy Commission, February.
- Faruqui, Ahmad and Melanie Mauldin. 2002. "The Barriers to Real Time Pricing: Separating Fact from Fiction," *Public Utilities Fortnightly*, July 15, pp. 30-40.
- Hirst, Eric. 2002(a). "Barriers to Price-Responsive Demand," prepared for Edison Electric Institute, June. Available at <http://www.ehirst.com/PDF/PRDBarriers.pdf>.
- \_\_\_\_\_. 2002(b) "The Financial and Physical Insurance Benefits of Price-Responsive Demand," *The Electricity Journal*, May, pp. 66-73.
- Patrick, Robert H. and Frank A. Wolak. 1997. "Estimating the Customer-Level Demand for Electricity Under Real-Time Market Prices," Stanford University working paper, available at <ftp://zia.stanford.edu/pub/papers/rtpap.pdf>.
- Ruff, Larry E. 2002(a). "Economic Principles of Demand Response in Electricity," prepared for Edison Electric Institute, October. Available at [http://www.eei.org/esg/policy/demand\\_response/economic\\_principles.pdf](http://www.eei.org/esg/policy/demand_response/economic_principles.pdf).
- \_\_\_\_\_. 2002(b). "Demand Response: Reality versus 'Resource' " *The Electricity Journal*, December, pp. 10-24.

## ENDNOTES

1. When this occurred in natural gas markets in the 1970s rationing was done in some areas by grandfathering all who already had gas service and requiring a queue for new hookups.
2. I assume here that capacity can be adjusted in very small increments, though the general conclusions hold even if investments are fairly lumpy.
3. Borenstein and Holland (2002) discusses in detail the efficiency of competitive equilibrium in this market and the social loss when some customers are charged flat rates.
4. Thus, the advance prices that retail sellers offer under these systems are not the equivalent of selling forward in a wholesale market, but rather are the equivalent of buyers holding infinite-quantity call options, though with the provision that the buyer cannot resell the commodity.
5. Braithwait and Eakin (2002) and Braithwait and O'Sheasy (2002) give overviews of the empirical evidence on the effectiveness of time-varying prices in eliciting demand response.
6. This comparison is not exactly the one we would like since real-time RTP prices are still announced slightly in advance, while the balancing market prices are determined during the actual period in which they are effective and are announced ex post. Thus, some of the 50%-55% that is not captured by day-ahead prices still wouldn't be captured by "real-time" RTP prices.
7. This is Pacific Gas & Electric's summer commercial TOU rate.
8. This number results from a regression, covering April 1998 to October 2000, of the hourly real-time wholesale price on dummy variables for each of the TOU periods (with observations weighted by quantity demanded, so the resulting TOU rates would exactly meet the retailers revenue requirements), where the coefficients on the TOU periods are allowed to differ in each summer/winter period. Comparing the explained sum of squares of such a regression to the explained sum of squares of a regression with just a single dummy for each summer/winter period gives the additional price variation captured by using the TOU periods rather than using a single constant price in each period. This

difference as a share of price variations was 6% in northern California and 13% in southern California. The comparable figure in PJM for the same period is 6%.

9. Customers that have TOU meters usually face on-peak and off-peak demand charges, with the on-peak maximum usage carrying a much higher price. Within the peak price TOU period (e.g., noon-6 p.m. on weekdays), however, the meters do not indicate at what time or day the maximum usage occurred.

10. There is another widespread view that many interruptible customers believed when they signed up that they were simply getting a price break and would never actually be called on to stop usage. Statements by these customers during the California electricity crisis of 2000-01 reinforced this view.

11. See Faruqui and George (2002) for a more detailed analysis of CPP.

12. In practice, either the utility makes a standard offer to all customers on the program or customers are asked to bid prices at which they would be willing to reduce their consumption.

13. See Ruff (2002a) for further discussion of baseline-setting problems.

14. This phenomenon, known as “adverse selection,” is a major concern in insurance markets, because high-risk individuals are more likely to buy insurance than low-risk.

15. Borenstein (2001) and Ruff(2002b) point out another problem with payments made under demand reduction programs. Many people argue that payments should be equal to the wholesale price, because the retailer saves this much by not buying the last kWh. But this overcompensates the customer, because the customer is also saving the retail price it would otherwise have to pay (and that the retailer does not receive if the customer reduces consumption). If the wholesale price is 30¢/kWh and the retail price is 8¢, then the appropriate payment for demand reduction would be 22¢/kWh not 30¢/kWh.

16. The rates might still be lower than under flat pricing since the reduced demand might lower the wholesale market price sufficiently to more than cover the cost of payments for demand reduction. Nonetheless, the revenue requirement implies that overall average rates will not be lower with demand reduction programs than with CPP.

17. See Faruqui, Hughes, and Mauldin (2002) and Faruqui and Mauldin (2002) for more discussion of two-part RTP programs.
18. Though the California economy was clearly entering a recession and the dot-com economy was declining, many companies still claimed that they were growing at phenomenal rates and therefore needed a CBL well above their past usage.
19. The spot price here could refer to a day-ahead price or a spot imbalance energy price.
20. In fact, the contract just has to have less variance than the spot price. It could, for instance, have a fuel adjustment clause.
21. This illustration, in which the long-term contract price is below the average spot price, is not meant to suggest that forward prices are systematically cheaper than spot prices. On average, the forward contract price will be about equal to the expected spot price during the life of the contract. Though it does not occur in this illustration, it is possible that this formula could result in negative prices in certain hours. This outcome could be easily avoided, however, with a small modification. A minimum price, say 1¢ per kWh, could be set and any resulting excess revenue could then be redistributed evenly among all other hours.
22. This illustration slightly overstates the monthly bill stability that could be achieved through 80% hedging because it assumes that the hedged quantity is 80% of the actual demand in each hour. The contract (or contracts) would quite likely hedge a larger quantity during periods when demand is anticipated to be high, but the variation would probably not match exactly the actual variation in consumption that occurs. Since price will be highest in periods when the quantity exceeds anticipated levels, the protection from the hedging contract would be slightly less than if it matched the actual consumption pattern exactly.
23. To be clear, the average price discussed here is the average over all customers on RTP, not a guaranteed average price for any one specific customer.
24. If the market is not competitive, then price is above the marginal cost of producing that unit and even the market price is giving too great an incentive to reduce consumption, leading to the standard deadweight loss outcome.

25. Ruff (2002b) makes a similar point about subsidizing demand reduction programs. Hirst (2002b) discusses the size of the cost premium from flat rate service.
26. It is quite possible that voluntary RTP would lower the cost of serving both groups, compared to no RTP, by reducing wholesale prices. Nonetheless, given any set of wholesale prices, the non-RTP participants will almost certainly be more costly to serve (on an average cost per kWh basis) than the RTP participants.
27. I abstract here from other costs, including T&D charges and the adders necessary to cover financial losses from the retailers long-term contracts or other stranded investments, all of which would be added to the charges of both groups. It is best to think of this approach as applying only to large industrial and commercial customers, all of whom could be fitted with real-time meters at relatively low cost.
28. By some, this has been called an attempt to prevent “cherry picking” that would drive up the cost to other customers. By others, this has been called an attempt to preserve the historical cross-subsidy.
29. Though many interruptible customers cannot actually be physically interrupted, as discussed above, the programs usually involve small number of large customers so it is feasible for the system operator to contact each by phone to see how they will respond to the call for demand reduction (and the extremely high retail price that accompanies the call).
30. I present an overview of the simulations and results here. For details, see Borenstein (2003).
31. I compress this two year dataset into one year by assuming that each load datum represents 30 minutes rather than one hour.
32. Here and throughout, the retail tariff is also assumed to include \$40/MWh for transmission and distribution.
33. See Borenstein (2000) for a numerical example of this.
34. Borenstein and Holland (2002) show that it is theoretically possible for capacity to increase or decrease with the introduction of RTP. These simulations demonstrate that for these reasonable parameters, the actual effect is a large decrease in capacity.

35. This is only an approximation, but it is accurate to within a few percentage points for looking at the response to price changes over the relevant range.

36. The increase in consumer surplus accrues to both customers on RTP and remaining flat-rate customers, though it goes disproportionately to customers on RTP. See Borenstein (2003).

37. See Michael Jaske's section of Borenstein, Jaske and Rosenfeld (2002). Also see Hirst (2002a) for a general discussion of the technological and bureaucratic barriers to RTP.

Table 1: Generation Costs Assumed in Long-Run RTP Simulations

Generation Type	Annual Fixed Cost	Variable Cost	Efficient Use of Capacity
Baseload	\$155,000/MW	\$15/MWh	operated > 4000 hrs/year
Mid-merit	\$75,000/MW	\$35/MWh	operated < 4000 and > 1000 hrs/year
Peaker	\$50,000/MW	\$60/MWh	operated < 1000 hrs/year

Table 2: Long-Run Effect of RTP compared to Flat Rate Tariff

**Capacity Changes**

Elasticity	Base Capacity (MW)	Mid-Merit Capacity (MW)	Peaker Capacity (MW)	Total Capacity (MW)
<i>Flat-Rate Tariff</i>				
—	27,491	5,472	12,912	45,875
<i>Real-Time Pricing</i>				
-0.025	27,631	5,257	7,798	40,686
-0.050	27,772	5,036	5,709	38,517
-0.075	27,912	4,822	4,281	37,015
-0.100	28,052	4,613	3,156	35,812
-0.150	28,331	4,188	1,446	33,965

---

**Annual Changes in Cost and Consumer Surplus***Flat-Rate Tariff*

Total Energy (MWh)	Total Energy Bill
236,796,579	\$9,265,850,141

*Real-Time Pricing*

Elasticity	Change in Total Energy (MWh)	Change in Total Energy Bill	Change in Consumer Surplus	$\Delta$ CS / Flat-Rate Energy Bill
-0.025	+800,441	-\$259,997,701	+\$233,825,698	+2.52%
-0.050	+1,462,755	-\$377,627,882	+\$335,913,373	+3.63%
-0.075	+2,097,715	-\$463,153,781	+\$409,994,816	+4.42%
-0.100	+2,723,226	-\$534,312,079	+\$472,647,808	+5.10%
-0.150	+3,968,346	-\$650,109,447	+\$578,088,940	+6.24%



## TITLES FOR FIGURES

Figure 1: Peak and Off-Peak Price With Fixed Capacity

Figure 2: Time-Invariant Pricing With Variable Demand

Figure 3: Excess Capacity Resulting From Time-Invariant Pricing

Figure 4: California Power Exchange Price versus System Load – June 2000

Figure 5: Real-Time Pricing With Monthly Bill Stability

(Assumes contract at 6 cents/kWh. Price include 4 cents/kWh T&D)









