

# Impacts of Responsive Load in PJM: Load Shifting and Real Time Pricing



## 1 Abstract

In PJM, 15% of electric generation capacity ran less than 96 hours, 1.1% of the time, over 2006. If retail prices reflected hourly wholesale market prices, customers would shift consumption away from peak hours and installed capacity could drop. I use PJM data to estimate consumer and producer savings from a change toward real-time pricing (RTP) or time-of-use (TOU) pricing. Surprisingly, neither RTP nor TOU has much effect on average price under plausible short-term consumer responses. Consumer plus producer surplus rises 2.8%-4.4% with RTP and 0.6%-1.0% with TOU. Peak capacity savings are seven times larger with RTP. Peak load drops by 10.4%-17.7% with RTP and only 1.1%-2.4% with TOU. Half of all possible customer savings from load shifting are obtained by shifting only 1.7% of all MWh to another time of day, indicating that only the largest customers need be responsive to get the majority of the short-run savings.

## 2 Introduction

The electricity industry uses much of its generation and transmission capacity only a small fraction of the time. Over the calendar year 2006, 15% of the generation capacity in the Pennsylvania-New Jersey-Maryland (PJM) territory ran less than 1.1% of the time (96 hours or less), and 20% of capacity ran less than 2.3% of the time (202 hours or less) [1]<sup>2</sup>. The result is tens of billions of dollars<sup>3</sup> invested in peaking generation that has low capital cost, but high generation cost and life cycle social cost.

The excessive peaking capacity has two causes. The first is technical: there must be enough system capacity to satisfy demand at all times or there will be a blackout. The second is regulatory: most customers pay a constant flat price for power rather than responding to the changing hourly price of the wholesale market. Flat-rate customers have no incentive to shift consumption away from times of peak demand. For example, a customer whose retail rate is \$0.10/kWh will pay the same price no matter whether the wholesale price reaches its limit of \$1/kWh during peak demand or drops to \$0/kWh during trough demand. If customers instead faced the changing wholesale price of

---

<sup>1</sup> Kathleen Spees is a doctoral student in Engineering and Public Policy at Carnegie Mellon University within the Carnegie Mellon Electricity Industry Center. Send correspondence to: Kathleen Spees, Carnegie Mellon University, EPP BPH 129, Pittsburgh, PA 15213; kspees@cmu.edu; 412.445.2694.

<sup>2</sup> This is based on the entire PJM hourly load profile in 2006 [1]. The system had 17.5% excess available generation capacity. I do not include generation excess at coincident peak load in this calculation because some generation excess is necessary for reliability purposes.

<sup>3</sup> At \$600/kW, a reasonable natural gas generator cost, this 15% of PJM's generation capacity is worth \$13 billion. At \$1800/kW, a reasonable price for a coal generator, 15% of PJM's capacity is worth \$39 billion.

electricity, as they do with most products including gasoline, natural gas, fruits, and vegetables, then they would buy less power at \$1/kWh and more at \$0/kWh. In doing so they would flatten demand over time, enabling society to diminish investments in peaking generators, instead increasing use of base-load units that have lower generation costs.

Some electricity customers face “time of use” (TOU) pricing that charges them a higher price during on-peak hours, with the fixed on-peak and off-peak rates calculated as the delivered cost averaged over a year. A few customers face “real time pricing” (RTP) where the hourly wholesale generation price determines the retail price. The TOU price gives better information and incentives than a single fixed tariff, but does not account for the times when wholesale prices spike because of high demand or equipment problems. Some view a TOU rate as a good compromise that frees customers from having to be informed about constantly changing prices and adjusting their consumption accordingly.

Few end users have any opportunity to react to real-time market conditions or to the location-specific costs of generation and transmission. A PJM survey of load-serving entities (LSE) reported that only 4.7% of end user MW are on rates directly or indirectly related to the real-time or day-ahead locational marginal price (LMP) [2, 3]<sup>4</sup>. Companies currently offering RTP rates usually have a variety of partial-hedging options as well [4]. Some additional customers are enrolled in direct load control, interruptible contracts, or other subsidy programs that offer curtailment incentives during the top few load hours per year. A Federal Energy Regulatory Committee (FERC) report estimates that 4% of peak MW in Reliability *First* Corporation (RFC) territory<sup>5</sup> could potentially have been curtailed via either RTP rates or non-price response programs, but the maximum response in 2005 was only 0.7% of MW [5]. Actual reductions are usually much smaller than program enrollments, partly because reduction is often voluntary [6].

I view the current flat tariff as both inefficient and inequitable. It is inefficient because it raises system costs and requires much more capital equipment to deliver the same quantity of power. It is inequitable, by my definition, because flat and counter-cyclical customers subsidize customers with high coincident peak demand.

I present a short-run analysis of a change to a more responsive demand-side market. In Section 5, I use one year of PJM data to build a supply model that implicitly accounts for dispatch constraints and varying conditions observed over a year. I use this model in three different simulations to estimate the impacts of responsive load. The first in Section 6 is an assumed load-shifting scenario that finds the effects of small changes in load profile on overall price. The load-shifting simulation does not consider customer time preference, but does show how quickly savings could be achieved. The final two simulations in Section 7 are more realistic; they use hourly demand curves to predict short-run impacts from change toward TOU or RTP from flat-rate pricing.

---

<sup>4</sup>. Estimate is from 3653 MW on locational marginal price (LMP) based rates and 69,063 MW represented in survey responses. I do not include load listed as switched to third-party suppliers in the calculation.

<sup>5</sup>. The RFC territory does not match up exactly with PJM territory.

### 3 Literature Review

Borenstein's long-run RTP analysis predicts more than double the peak load savings I predict in my short-run analysis, see Section 7 [7]. His conclusion results from using a long-term supply curve in estimating hourly equilibrium conditions. Because Borenstein includes capital costs in his supply curves, he predicts hourly prices up to \$90,772/MWh; this implies that customers could spend 22% of the yearly bill in one peak demand hour. Those high prices would only be possible if market rules change dramatically since hourly prices are hard-capped at \$1000/MWh<sup>6</sup> in all but one United States market and determined based on short-run conditions with a fixed generation portfolio [9]. Further, I believe that Borenstein's exercise is intended to be primarily illustrative on peak load reductions since his resulting load duration curves are abruptly leveled off on the high end. My short-run analysis reflects current PJM conditions because I use observed market data.

Holland and Mansur predict less than half the short-term peak load savings that I predict from RTP, see Section 7 [10, 11]. The modest impact is due to their method of using one constant stacked marginal cost curve to represent supply over the entire year<sup>7</sup>. I use observed market prices to account for transmission and other constraints<sup>8</sup> while they assume constraint-free economic dispatch of system generators to estimate marginal cost. Holland and Mansur attempt to correct for one of these constraints, generator availability, by discounting the capacity of each generator by an expected "outage" factor, but the method cannot capture the observed phenomenon of very high prices at moderate demand levels. Based on my own empirical analysis, I find that a constraint-free stacked marginal cost curve underestimates price by \$15.88/MWh on average<sup>9</sup>, and, more importantly, it also underestimates the *slope* of the real supply curve. The supply curve slope determines the impact that a small change in load has on price, meaning that ignoring transmission and dispatch constraints can lead to qualitatively wrong policy conclusions for RTP. For an empirical comparison of observed prices to constraint-free dispatch curves, please contact the author.

Power engineers account for real-time transmission constraints by solving the security-constrained direct-current optimal power flow (DCOPF) problem in example cases. This approach is similar to how PJM sets market prices. Wang, Redondo, and Galiana used a DCOPF-based model to examine demand-side participation in wholesale energy and

---

<sup>6</sup>. California ISO is the exception with a \$400/MWh soft cap on energy and ancillary service bids [8]. Generators may bid above a soft price cap and will be paid as bid; other generators will receive payment only as high as the cap. The neighboring Western Electricity Coordinating Council (WECC) has the same price caps although WECC is not a market operator.

<sup>7</sup>. Their stacked marginal cost curve is based on generator heat rates, fuel prices, emissions prices, and other publicly available data for the time frame in question.

<sup>8</sup>. Examples of other constraints include limits on run times, ramp rates, reserve margins, local reactive power generation, scheduled maintenance etc.

<sup>9</sup>. The estimate uses PJM generator bid data over a calendar year from June 2005 through May 2006. The bid data are publicly available after a six-month delay. Full details of this calculation are available in **Error! Reference source not found.**

ancillary services markets [12]. Their results indicate that demand participation erodes generator market power. However, results from test systems with a few buses do not translate directly into implications for the PJM system with roughly 7800 pricing points. Fitting supply curves to daily market data incorporates these constraints.

## 4 Data

My data are system-wide hourly PJM market clearing results. I examine aggregate load and PJM average prices<sup>10</sup> in the day-ahead and real-time markets over 2006 [1]. Day-ahead demand bids  $L_{DA}$  from LSEs are charged at the day-ahead price  $P_{DA}$ , the real-time increment or decrement  $L_{RT}-L_{DA}$  is charged or credited at the real-time price  $P_{RT}$ . Overall revenue and price are calculated in Equations (1) and (2).

$$R = L_{DA} \cdot P_{DA} + (L_{RT} - L_{DA}) \cdot P_{RT} \quad (1)$$

$$P_0 = \frac{\sum_{hours} R}{\sum_{hours} L_{RT}} \quad (2)$$

Overall realized price and the real-time demand for each hour are the most accurate data for evaluating demand response. In the implementation of RTP rates, customers should have access to both day-ahead and real-time market prices. I assume that nearly all power continues to be purchased in the day-ahead market; both markets are counted as RTP.

## 5 Market Model

I construct a short-term equilibrium model accounting for producer, consumer, and local utility participation. Results from the full model for RTP and TOU pricing are in Section 7. The load-shifting scenario in Section 6 uses only the supply-side model developed in this section.

### 5.1 Model

My base case model treats the retail and wholesale markets separately as shown in Figure 1. In the retail market, I assume that all consumers currently pay a flat rate  $P_0$  for all their power, making the supply curve appear completely elastic to consumers. In the wholesale market, the market operator treats hourly demand  $L_0$  as completely unresponsive to price. While each hour has wholesale price  $P_W$  above or below retail price, the profits and losses are temporarily absorbed by the LSE and sum to zero over

<sup>10</sup> The PJM price is a load-weighted average of all system LMPs.

the year. This disconnect between wholesale and retail is a good characterization of current conditions since few customers face RTP [2, 3, 5, 13].

Under TOU the retail price takes on a value of  $p_{on}$  during on-peak hours and  $p_{off}$  during off-peak hours. In PJM off-peak hours are weeknights 11 PM to 7 AM and all day on weekends and the six NERC holidays [14]. On and off-peak prices are set so that local utility profit sums to zero over on-peak hours and off-peak hours separately.

When I model RTP, I set the retail price equal to the wholesale price, eliminating the disconnect between wholesale and retail (I neglect distribution costs).

### 3.1

I assume that each hour has a unique demand curve with constant elasticity as shown in Equation (3) where the hourly parameters  $\beta$  are determined by base case price and hourly load [7, 10].

$$P_D(L) = \beta \cdot L^{1/E} \tag{3}$$

$$\beta = \frac{P_0}{L_0^{1/E}}$$

The right side of (3) is replaced with the retail price  $P_D(L)$  that applies in the flat (4), TOU (5), or RTP (6) cases.

$$P_D(L) = p_0 \tag{4}$$

$$P_D(L) = P_{TOU} = \begin{cases} p_{on} \\ p_{off} \end{cases} \tag{5}$$

$$P_D(L) = P_S(L) \tag{6}$$

In order to model demand using the most realistic elasticities, I use estimates from the literature. A 1984<sup>11</sup> review of 34 studies found short run and long run price elasticities to be approximately -0.20 and -0.90 respectively, implying that a 10% price increase would reduce consumption by 2% in the short-run and 9% in the long-run [13]. Most of these estimates were made based on a change from one flat rate for power to another, not responses to hourly changing prices, and so the short run number only hints at the appropriate number for my purposes.

More telling is that after 5 years of experience with default RTP for customers larger than 2 MW, Niagara-Mohawk Power Corporation has observed an average demand elasticity of substitution of -0.11 [15, 16]. A Department of Energy study reviewed price

<sup>11</sup>. The short run numbers were recently updated in another review of 36 estimates with a median of -0.28.

elasticities of substitution under TOU, critical peak pricing (CPP), and day-ahead RTP situations [17]. The range of elasticities of substitution was 0.02 to 0.27.

The level of responsiveness that would be observed under RTP is uncertain and could depend on a variety of factors including customer class, weather, and enabling technology. Regardless, there have been enough empirical estimates to place the plausible short-run elasticities of demand between 0 and -0.4 under RTP conditions. I examine this full range. I will not specify exactly how an aggregate elasticity is achieved, for example having all customers on RTP with an elasticity of -0.1 would be approximately the same as having only half of all MW on RTP with an elasticity -0.2.

### 3

At one extreme, I might hypothesize that the wholesale supply-side relationship between price and load is the same over an entire year. At the other extreme, I might hypothesize that the relationship is unique to each day. The market clearing price at a specified load level may differ from one day to another because some generating units or transmission lines are not available, fuel prices have changed, or weather is impeding supply. Fitting unique parameters for each day would give a better fit than insisting that one set of parameters must fit the entire year. However, the former is not a parsimonious model and says nothing about what parameter values should be used in future days.

The wholesale price of electricity for each hour in a day follows a predictable pattern of being low in the early morning and at night with one or two peaks during the day. I fit the price and load data for each day with a third-degree polynomial. To investigate the similarity of the polynomial parameters across days, I employ dummy variables, taking on values of 0 or 1.

Equation (7) models price as a function of load represented by an intercept, load, load squared, and load cubed. The equation uses dummy variables  $\delta_l$  and  $\delta_o$  to allow for the possibility that the coefficient of load and the intercept might vary each day. I also examined the possibility that the coefficients of the squared and cubed terms take on unique values each day but determined that the additional dummy variables improved explanatory power very little. I selected (7) as a model with good explanatory power, only half the number of parameters as employing the additional two dummy variables, and as a good fit to the plotted data. For detailed results from trying a range of models, please contact the author.

$$P_s(L) = a \cdot L^3 + b \cdot L^2 + \sum_{t=1}^n \{ \delta_l \cdot c_t \cdot L + \delta_o \cdot d_t \} \quad (7)$$

The adjusted  $R^2$  is 0.949, the F-statistic of 223 is highly significant<sup>12</sup>, and the estimated parameters  $a$  and  $b$  are highly significant<sup>13</sup> all with p-values  $\square$  0.001.

<sup>12</sup>. Model significance test has  $F(731,8028) = 223$  with p-value  $\square$  0.001.



Changes in consumer surplus  $\Delta CS$  and producer surplus  $\Delta PS$  between flat rate and RTP conditions are calculated in Equations (8) and (9) and shown graphically in Figure 1. Producer surplus is easier to calculate by integrating over load than over price. Change in consumer surplus in Equation (8) can be calculated in the TOU case by replacing  $P^*$  with the retail TOU price  $p_{off}$  or  $p_{on}$ . Change in producer surplus calculated in Equation (9) is the same formula under a change toward TOU or RTP because the wholesale electric price determines the producer surplus.

$$\Delta CS = \sum_{hours} \int_{P^*}^{P_0} L(P_D) \partial P = \sum_{hours} \int_{P^*}^{P_0} \left( \frac{P_D}{\beta} \right)^E \partial P = \sum_{hours} \left( \frac{1}{E+1} \right) \left( \frac{P_D}{\beta} \right)^{E+1} \Big|_{P^*}^{P_0} \quad (8)$$

$$\begin{aligned} \Delta PS &= \sum_{hours} \int_{P_S(L_0)}^{P^*} L(P_S) \partial P = \sum_{hours} \left\{ P^* L^* - P_0 L_0 - \int_{L_0}^{L^*} P_S(L) \partial L \right\} \\ \Delta PS &= \sum_{hours} \left\{ P^* L^* - P_0 L_0 - \int_{L_0}^{L^*} (aL^3 + bL^2 + cL + d) \partial L \right\} \\ \Delta PS &= \sum_{hours} \left\{ P^* L^* - P_0 L_0 - \left[ \left( \frac{a}{4} L^4 + \frac{b}{3} L^3 + \frac{c}{2} L^2 + dL \right) \right]_{L_0}^{L^*} \right\} \end{aligned} \quad (9)$$

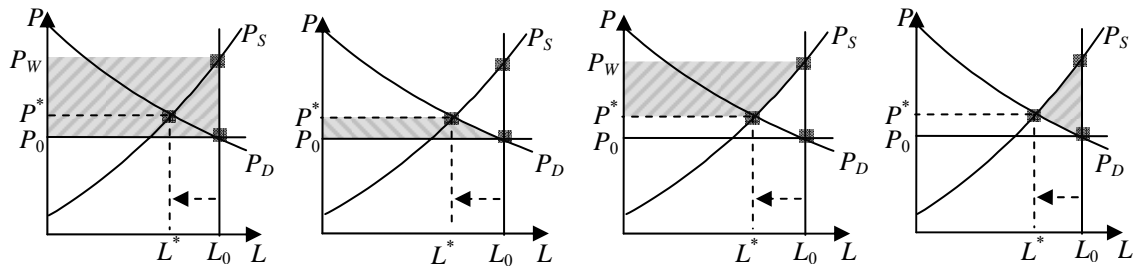
With flat-rate or TOU pricing there is deadweight loss in both high-priced hours and low-priced hours. Because the RTP case has no deadweight loss, I calculate the deadweight loss in the flat rate and TOU cases based on the surplus changes in Equation (10). Both deadweight loss and LSE profit  $\Pi$  are shown in Figure 1 for a sample high-priced hour.

$$\begin{aligned} DW_{flat} &= \Delta \Pi_{flat}^{RTP} + \Delta CS_{flat}^{RTP} + \Delta PS_{flat}^{RTP} = \Delta CS_{flat}^{RTP} + \Delta PS_{flat}^{RTP} \\ DW_{TOU} &= DW_{flat} - \Delta DW_{flat}^{TOU} = DW_{flat} - (\Delta CS_{flat}^{TOU} + \Delta PS_{flat}^{TOU}) \end{aligned} \quad (10)$$

The model is shown graphically in Figure 1 for a sample high-priced hour. The base-case retail market is represented by demand curve  $P_D(L)$  and completely elastic supply  $P_0$ ; the wholesale market is represented by supply curve  $P_S(L)$  and completely inelastic demand  $L_0$ . The base case model has two different resulting prices  $P_0$  and  $P_w$  that apply in the retail and wholesale markets respectively, but resulting load has to be the same in both. The arrow shows how load drops under RTP when the integrated market is representing by supply and demand curves  $P_D(L)$  and  $P_S(L)$ .

<sup>13</sup>. Studentized t-test have  $ta(8028) = 10.9$  and  $tb(8028) = 33.0$  with p-values  $\square 0.001$  in each case.

The shaded areas in Figure 1 show from left to right: the LSE's hourly deficit under flat-rate pricing, hourly consumer surplus drop in moving from flat-rate to RTP as in (8), hourly consumer surplus drop in moving from flat-rate to RTP as in (9), and hourly deadweight loss as in (10). Note that in a corollary low-price hour, load would increase under RTP but the mathematical definitions would hold. Although the local utility may have a positive or negative profit in any one hour with TOU or flat rate, it has zero profit over the year under any of these pricing scenarios.



**Figure 1.** From left: loss to the local utility under a flat rate since payments to wholesale exceeded revenues from customers, drop in consumer surplus moving from flat rates to RTP, drop in producer surplus moving from flat rates to RTP, and deadweight loss under a flat rate. Note that this is the case of high wholesale prices; there are corollary hours when the wholesale price under flat rates is lower than the retail price.

## 6 Load Shifting

Assume that customers can be induced to shift their demand to be more level over the day. Although the resulting load profiles may not be realistic, I use this simulation to show how much shifting is necessary to flatten load and how quickly savings can be achieved.

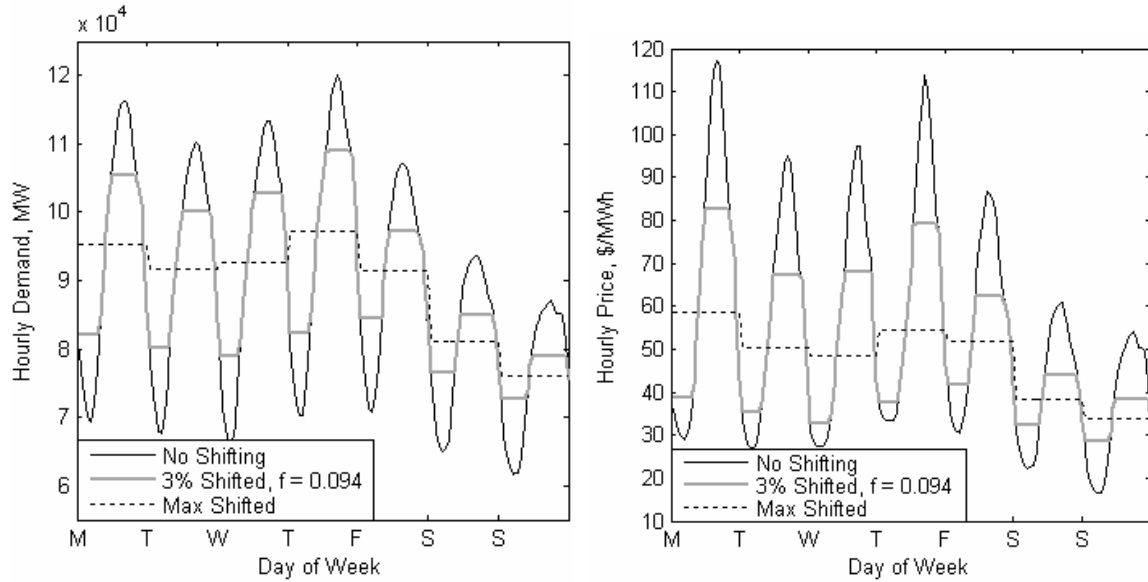
### 6 M

I scale possible consumer savings from demand response by incrementally shifting load to achieve a totally flat daily load profile without changing total consumption. Although this method does not consider real-world preference effects, it does set an upper bound on customer savings. The simulation allows load shifting to any other time of day but does not allow shifting from one day to another.

For a particular day, I simulate shifting an increment of demand from the highest load hour to the lowest load hour. I continue shifting demand increments so that there is one wholesale price for the hours of greatest use and another (lower) wholesale price for the hours of least use. The maximum fraction  $f$  that is curtailed off the peak load hours is the same for all days. I stop shifting load when the quantity and wholesale price are the same for the high and low-priced hours. The simulation reaches maximum shifting with 5.3% of all MWh shifted away from peak hours and  $f = 0.158$  (or 15.8% of MW) at which point the load profile is flat over each day, but not between days.



Figure 2 illustrates the effects of shifting on load and price profiles of one week beginning Monday, June 19, 2006. This week originally exhibited moderately high load and price. Results are shown when 3% of all yearly MWh are shifted and after the maximum shifting of 5.3% of all yearly MWh. This method does not change total daily consumption in MWh, but the extremes of usage and price variation are reduced.

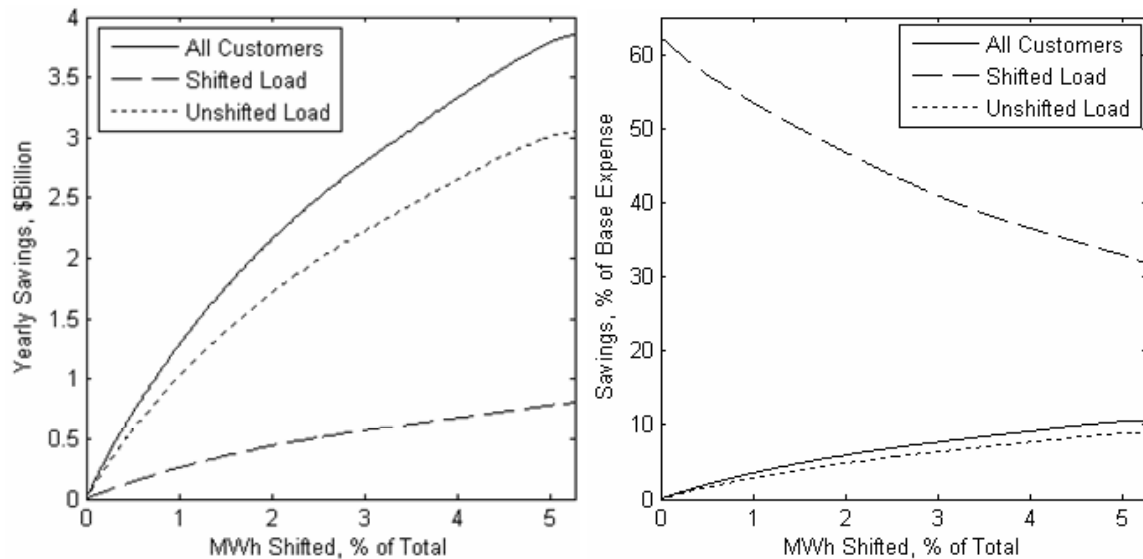


**Figure 2.** Load and price profiles for a July week; base case, 3% shifting ( $f = 0.093$ ), and maximum shift.

## 6 R

I remind the reader that the load shifts are imposed, rather than resulting from consumer preferences and so no conclusions can be drawn about consumers being better or worse off.

Customer expenditure savings from load shifting are shown in Figure 3. Savings are also split out by the amounts received by shifters and those received by free riders that do nothing. The left-hand plot in Figure 3 displays decreasing marginal savings with more shifting; when the daily load is leveled, there are no further savings. The right-hand plot of Figure 3 shows that shifters' percentage savings drop with increased shifting. This is because the price differential over a given day can be large under current conditions but approaches zero in the limit; small marginal savings steadily reduce average calculated savings. Total customer savings increase with the amount of shifting with an ultimate limit of 10.7% of the annual electric bill.



**Figure 3.** Savings to shifters, free riders, and total in dollars (left) and as a percentage of bill (right).

Load shifting reduces peak load dramatically as shown in Table 1, obviating the need for costly investment in generation and transmission.

**Table 1.** Peak load and overall cost savings with daily shifting.

Shifted Load, %	Peak Load, GW	Peak Load Saved	Total Expense, \$Billion	Average Cost, \$/MWh	Customer Bill Savings
0%	145	0.0%	\$36.17	\$51.96	0.0%
1%	138	4.8%	\$34.90	\$50.13	3.5%
2%	134	7.3%	\$34.03	\$48.88	5.9%
3%	131	9.3%	\$33.37	\$47.94	7.7%
4%	128	11.6%	\$32.84	\$47.17	9.2%
5%	122	15.8%	\$32.38	\$46.51	10.5%
5.3%	122	15.8%	\$32.32	\$46.43	10.7%

Table 2 shows how quickly customer savings are reached by load shifting. Half of all the possible savings from load shifting are achieved by shifting only 1.69% of all energy. This indicates that a small amount of demand response is all that is needed to get most of the benefits.

**Table 2.** Load shifting necessary to achieve a portion of limiting savings with daily shifting.

<b>% of Savings in Limit</b>	<b>% Load Shifted</b>	<b>Maximum Hourly % Curtailed</b>
25%	0.70%	3.9%
50%	1.69%	6.6%
75%	3.15%	9.6%
90%	4.26%	12.4%
95%	4.66%	14.0%
99%	5.06%	16.5%

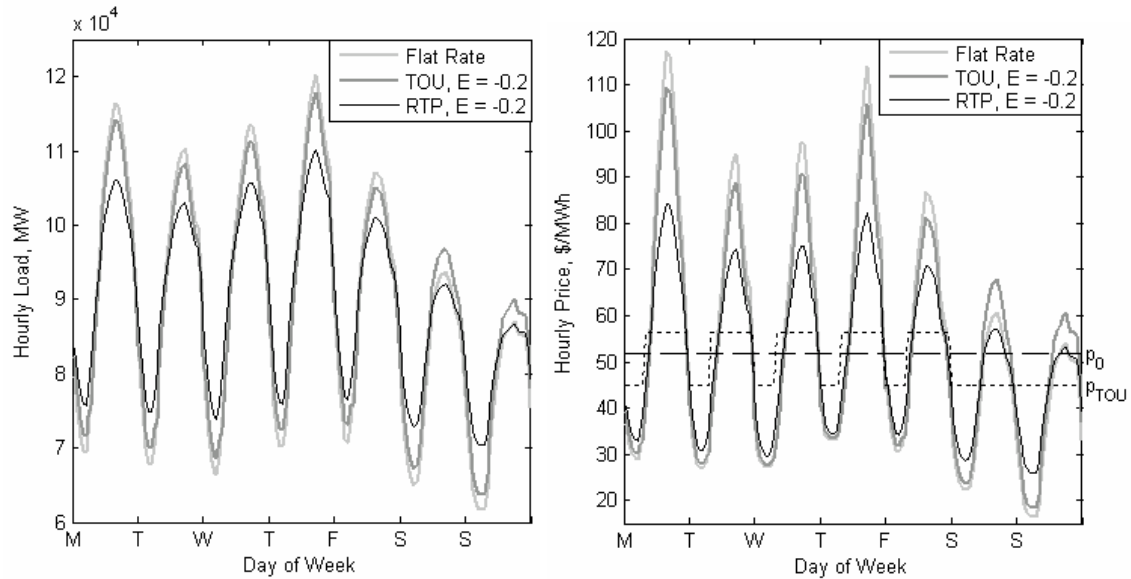
## 7 Time of Use and Real Time Pricing

I turn from calculating the savings from assuming that load can be shifted to an analysis of how much consumers would shift load in response to price differentiates between high and low demand hours. I use a simulation to determine the magnitude of effects from a change to RTP or TOU.

### 1 Introduction

The new price and load under RTP and TOU conditions are calculated as in Section 5. Figure 4 shows load and wholesale price profiles  $P_S$  over a week in the base case, under TOU, and under RTP conditions with elasticity -0.2. Under RTP, the price that consumers face is the same as the one paid to generators in the wholesale market,  $P_D = P_S$  as in (6). Under flat or TOU rates, the wholesale price  $P_S$  can be higher or lower than the retail prices. For reference the flat and TOU retail rates  $p_0$  and  $p_{TOU}$  are shown in dashed lines for the flat-rate and TOU cases respectively. The June week shown originally had moderately high load and wholesale price, so the RTP case shows steep drops in price and load during peak hours.

The left-hand graph in Figure 4 shows that RTP reduces peak loads much more than TOU pricing, which is only slightly better than flat rate pricing. The right-hand graph shows wholesale prices reflecting the marginal generation cost as solid lines; retail tariffs are in dashed lines. Under RTP the wholesale and retail prices are the same solid line. Wholesale price peaks are moderated much more under RTP than under TOU pricing. A TOU rate actually exacerbates wholesale price peaks on weekends because end users see the off-peak price all day.



**Figure 4.** Load and price profiles with elasticity -0.2 for a July week with flat rates, TOU, and RTP. Retail prices are shown in dashed lines while wholesale prices are in solid lines, except for the RTP case in which retail prices equal wholesale prices.

## 2

Market outcomes depend on the assumed demand elasticity<sup>14</sup>. Table 3 and Table 4 summarize impacts on consumption, expense, average price, and peak load with TOU and RTP rates respectively. The impacts from TOU pricing are a fraction of those from RTP. Impacts from TOU in peak load shaved, consumption increase, and consumer expense saved are never more than 14.4%, 22.3%, and 21.9% respectively of the impacts from changing to RTP at any elasticity.

<sup>14</sup>. Customers are more responsive when elasticity is more negative; responsiveness increases as one moves to the left in these plots.

Impacts on consumer expense and consumption increase are small under either rate structure change. The most striking result in these tables is that with RTP, peak load reductions are large even with highly (but not completely) inelastic demand. I estimate a 10.4% reduction in peak demand at elasticity  $E = -0.1$ , a huge reduction at a modest assumed responsiveness. Holland and Mansur's prediction with all customers on RTP at this same elasticity is less than half ours at 3.91%, while Borenstein's estimate is more than twice the size at 24.5%<sup>15</sup> [7, 10].

**Table 3.** Load increase, peak shaving, and price savings with TOU pricing.

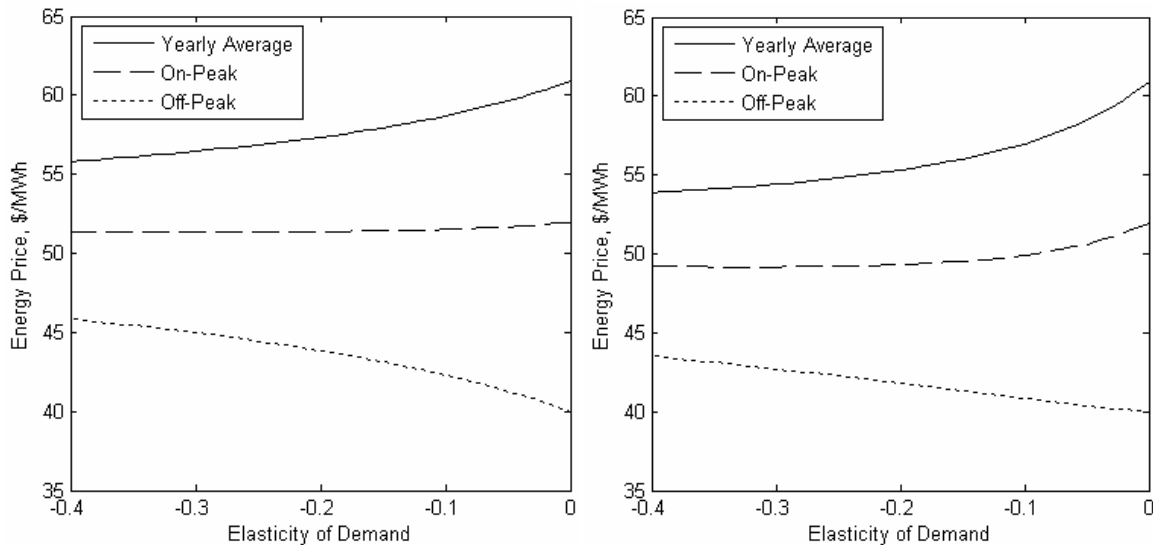
Elasticity of Demand	Peak Load, GW	Peak Load Saved	Total Energy, TWh	Consumption Increase	Total Expense, \$Billion	Consumer Expense Saved	Average Price, \$/MWh
0	145	0.00%	696	0.00%	\$36.17	0.00%	\$51.96
-0.05	144	0.57%	697	0.08%	\$36.04	0.38%	\$51.72
-0.1	143	1.12%	697	0.18%	\$35.95	0.62%	\$51.54
-0.15	143	1.55%	698	0.27%	\$35.91	0.73%	\$51.44
-0.2	142	1.89%	699	0.35%	\$35.90	0.76%	\$51.38
-0.25	142	2.18%	699	0.42%	\$35.90	0.75%	\$51.35
-0.3	141	2.42%	700	0.48%	\$35.91	0.71%	\$51.34
-0.35	141	2.62%	700	0.54%	\$35.93	0.67%	\$51.34
-0.4	141	2.79%	700	0.58%	\$35.95	0.61%	\$51.34

**Table 4.** Load increase, peak shaving, and price savings with RTP.

Elasticity of Demand	Peak Load, GW	Peak Load Saved	Total Energy, TWh	Consumption Increase	Total Expense, \$Billion	Consumer Expense Saved	Average Price, \$/MWh
0	145	0.0%	696	0.00%	\$36.17	0.00%	\$51.96
-0.05	137	5.7%	699	0.38%	\$35.52	1.82%	\$50.82
-0.1	130	10.4%	702	0.83%	\$35.11	2.93%	\$50.02
-0.15	126	13.3%	705	1.23%	\$34.94	3.39%	\$49.59
-0.2	123	15.1%	707	1.59%	\$34.90	3.51%	\$49.35
-0.25	121	16.6%	709	1.91%	\$34.93	3.44%	\$49.23
-0.3	119	17.7%	711	2.20%	\$34.99	3.27%	\$49.18
-0.35	118	18.7%	713	2.44%	\$35.07	3.04%	\$49.18
-0.4	117	19.5%	715	2.67%	\$35.16	2.80%	\$49.20

<sup>15</sup> Holland and Mansur also predict a 5.88% peak load reduction at  $E = -0.2$ , where I predict a 15.1% savings. Borenstein also predicts 35.2% peak load reduction at  $E = -0.3$  where I predict a 17.7% savings. The modest impacts predicted by Holland and Mansur are largely dictated by their method of using a stacked bid curve, please contact the author. Borenstein's large projected peak reduction has to be understood knowing that his supply curve comprised of three generator types results in a load duration curve that is completely chopped off on the high end; he does not argue that this is a realistic resulting load duration curve.

On-peak, off-peak, and average wholesale prices are shown in the left-hand side of Figure 5 for TOU pricing and in the right-hand side for RTP. Prices drop more with RTP; they are about 4% lower. Both schemes moderate on-peak and off-peak prices on average. Table 5 shows the same on- and off-peak prices as in Figure 5 at sample customer elasticities as well as showing results for the most extreme prices. A regulator looking only at prices might be deceived by the apparently small difference between RTP and TOU on average prices.



**Figure 5.** On-peak, off-peak, and average prices under the TOU scenario (left) and RTP scenario (right).

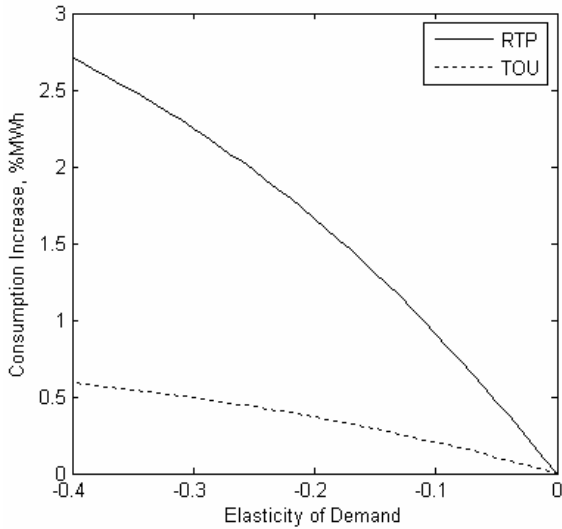
**Table 5.** Yearly prices with a change to TOU or RTP.

Elasticity of Demand	On-Peak Price, \$/MWh		Off-Peak Price, \$/MWh		Highest Price, \$/MWh		Lowest Price, \$/MWh	
	TOU	RTP	TOU	RTP	TOU	RTP	TOU	RTP
0	\$60.92	\$60.92	\$40.01	\$40.01	\$292.03	\$292.03	\$5.23	\$5.23
-0.05	\$59.87	\$58.86	\$41.03	\$40.28	\$286.12	\$234.09	\$5.89	\$10.13
-0.1	\$58.86	\$57.23	\$42.08	\$40.72	\$280.43	\$189.03	\$6.58	\$14.42
-0.15	\$58.08	\$56.17	\$42.96	\$41.20	\$276.02	\$160.92	\$7.13	\$17.76
-0.2	\$57.46	\$55.43	\$43.70	\$41.69	\$272.50	\$141.83	\$7.60	\$20.50
-0.25	\$56.95	\$54.89	\$44.33	\$42.16	\$269.63	\$128.14	\$7.99	\$22.80
-0.3	\$56.53	\$54.49	\$44.87	\$42.61	\$267.23	\$117.90	\$8.33	\$24.78
-0.35	\$56.18	\$54.17	\$45.34	\$43.04	\$265.20	\$109.99	\$8.62	\$25.96
-0.4	\$55.87	\$53.92	\$45.75	\$43.43	\$263.46	\$103.72	\$8.88	\$26.86

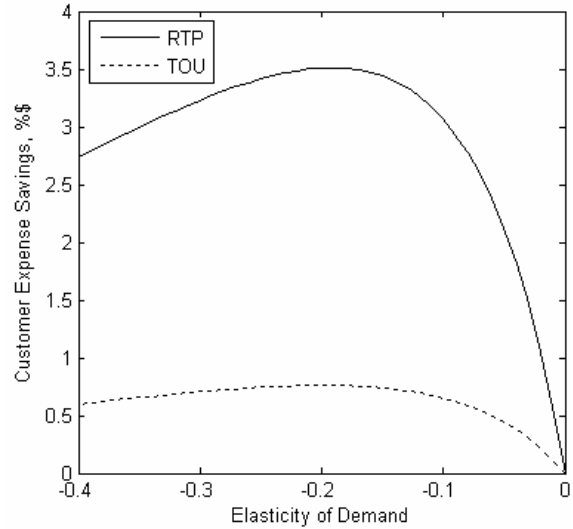
Consumers elect to buy more energy under RTP or TOU conditions as shown in Figure 6. Note that TOU and RTP result in prices above the flat-rate price for some hours and below it for others. The result is a drop in the quantity demanded during the high price period and an increase during the low price period. Since there are net customer savings, there is a small net increase in the quantity demanded. Marginal impacts diminish with more responsive load. Customer expenditure on electricity decreases steeply if elasticity is low in magnitude as shown in Figure 7. With inelastic demand most of the changes in

consumption patterns are small reductions at peak prices. With greater elasticity, dollar savings drop as the effect of greater consumption dominates the overall expense.

These RTP results are explained by the large positive skew in electricity prices and the increasing steepness of supply curves at high load. Large price reductions from small amounts of curtailment at high prices dominate results at elasticities near zero. With increasing responsiveness, the load profile becomes flatter and flatter but overall consumption increases. Under these conditions, the effect of the consumption increase dominates other results. Results with TOU pricing have similar characteristics but only a fraction of the magnitude.



**Figure 6.** Consumption increase, TOU and RTP.



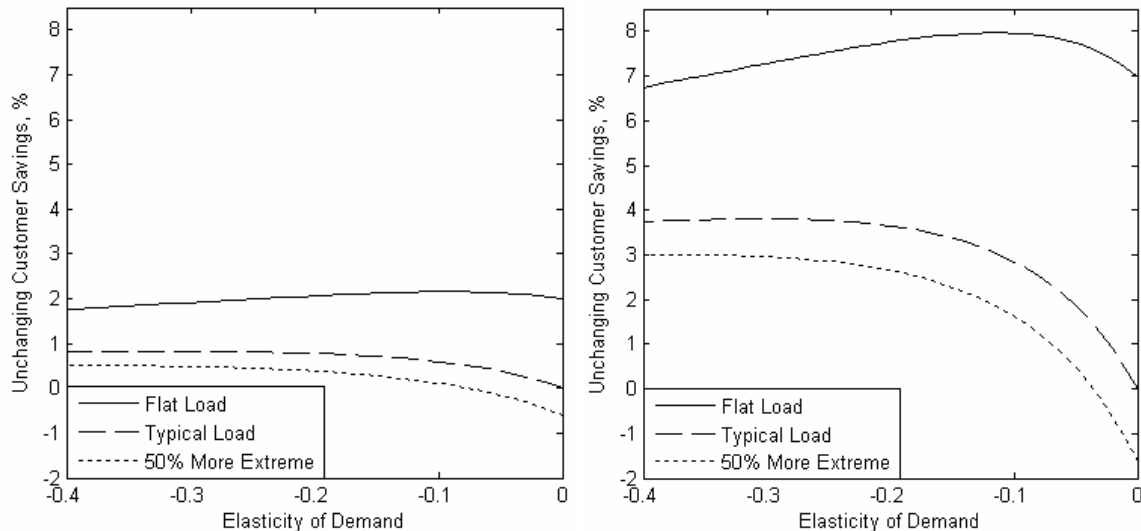
**Figure 7.** Customer bill savings, TOU and RTP.

Because consumers are buying more energy with less total expenditure, the overall impact on consumers is more easily understood by looking at a customer who refuses to change behavior as others do under TOU or RTP. In Figure 8, savings are shown for a single customer who has elasticity zero, while the aggregate system has an elasticity shown on the x-axis. I show savings for three types of customers:

- Flat* – Customer uses a constant level of power during all hours of the year.
- Typical* – Customer load profile is proportional to the original system load profile.
- 50% More Extreme* – During each hour, the customer demands the typical customer’s load *plus* an additional 50% of the difference between the typical customer’s load for that hour and the minimum load for the day.

An unchanging typical customer saves less per unit than a responsive customer, but slightly more overall because she does not increase consumption<sup>16</sup>. More interesting is that a flat customer would save 7.0% of her annual electric bill even if *no one* responded to price. She would save the amount that currently goes to subsidize the excesses of more peaky customers. This savings highlights the issue of equity that I raised earlier: under flat rates, moderate and counter-cyclical customers subsidize the consumption of customers with high coincident peak loads.

The more extreme customer loses money under RTP if no one responds, but will have net savings if the aggregate elasticity is even slightly responsive,  $E \leq -0.04$ .



**Figure 8.** Expense savings to an unresponsive customer when others respond, TOU (left) and RTP (right).

<sup>16</sup>. At  $E = -0.2$ , the typical responsive customer saves 5.0% per unit and 3.5% overall; the typical unresponsive customer saves 3.6% although her quantity consumed is constant.

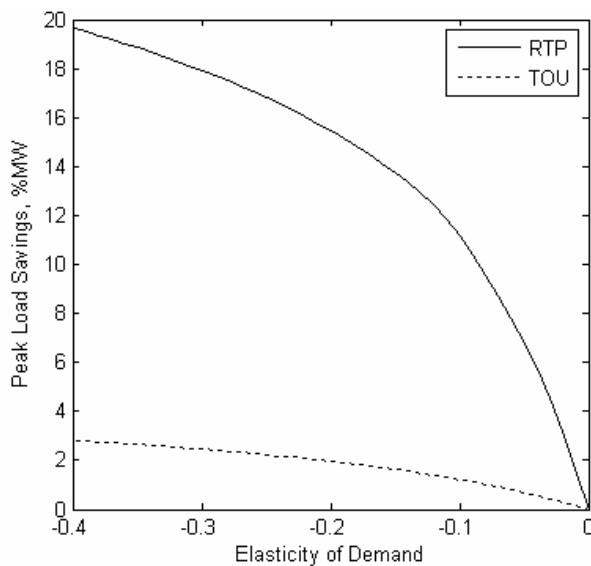


Peak load reductions are extreme with a small amount of responsiveness but marginal savings taper with greater responsiveness as shown in Figure 9. Discontinuities in Figure 9 are caused by a change in the day upon which peak load is observed.

The large peak load savings under RTP have huge implications for the total system cost. Peak load determines the total capacity investment necessary for the system to operate reliably. Although no savings will be made on peak capacity that has already been built, there will be savings via unneeded capacity investment as generators have to be replaced or load increases over time. At elasticity -0.2, peak load drops by 15.1% with RTP. At that level, an overnight capacity value of \$600/kW or \$1800/kW, corresponding roughly with the overnight capital costs of gas and coal generation, translates into a dollar savings of \$13- \$39 billion from a change to RTP. A change to TOU pricing would reduce \$1.7 to \$5.0 billion in capacity investments under the same conditions.

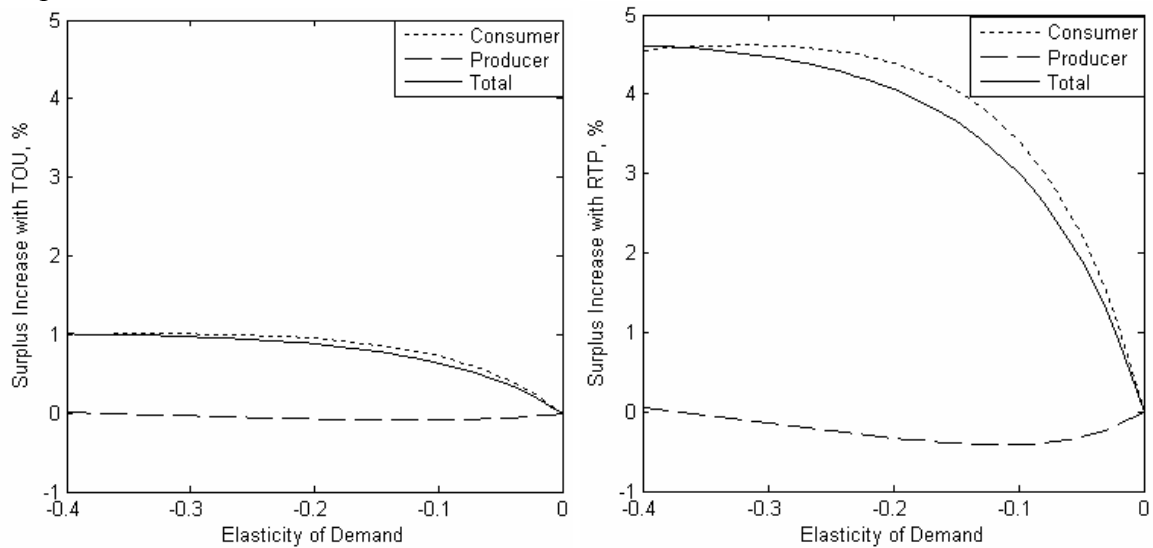
If state regulators and utilities begin to treat RTP as an alternative to investments in new generating capacity, then they will have to compare the costs of investing in new capacity against the costs of implementing RTP. At \$13 billion in avoided capacity costs, an integrated resources planner would be willing to spend \$257 for each of the 51 million people in PJM territory to implement RTP [18]. Compared to the hardware and installation costs of \$123-\$215 per unit for the advanced metering infrastructure required to implement RTP, these capacity savings justify RTP rates starting with the largest and most responsive customers [5].

I conjecture that only large customers need to face RTP to achieve most of these savings. From the experience in Niagara Mohawk, the “18% [of customers] with elasticities greater than -0.1 provide 85% of the aggregate price response” [15]. If only a fraction of customers need smart meters, and automatic energy managers to respond to RTPs, then the cost of implementing RTP would be much smaller than the social benefit, with all customers receiving some benefits via lower average price.



**Figure 9.** Peak load reductions, TOU and RTP.

Figure 10 and Table 6 show surplus increases with a time-varying rate. Neither consumer nor producer surplus changes monotonically with elasticity. Producer surplus drops slightly with peak price reductions but then increases with overall consumption. Producer surplus is equal to revenue minus operating costs and so indicates profitability if capital costs are not considered. Because I see almost no change in producer surplus, these results indicate that producers will not see the large reduction in profits that they might have feared from RTP. There is no change in consumer surplus for an elasticity of zero, but for an elasticity of -0.2, consumer surplus increases 0.7% for TOU pricing and 3.2% for RTP. I find that TOU pricing has only 20.3%-21.8% the impact in increasing total surplus that RTP would have<sup>17</sup>. No matter what the assumed elasticity, consumer surplus increases with RTP or TOU<sup>18</sup>.



**Figure 10.** Surplus increases with TOU (left) and RTP (right) as a percent of baseline expense.

**Table 6.** Economic outcomes with RTP as a percentage of baseline expenditure.

Elasticity of Demand	Flat-Rate Deadweight Loss	TOU Rate Deadweight Loss	Surplus Increase with TOU			Surplus Increase with RTP		
			Consumer	Producer	Total	Consumer	Producer	Total
0	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
-0.05	1.61%	1.28%	0.39%	-0.06%	0.33%	1.89%	-0.28%	1.61%
-0.1	2.82%	2.22%	0.68%	-0.09%	0.60%	3.22%	-0.40%	2.82%
-0.15	3.54%	2.78%	0.85%	-0.09%	0.76%	3.95%	-0.40%	3.54%
-0.2	3.99%	3.13%	0.94%	-0.08%	0.86%	4.34%	-0.34%	3.99%
-0.25	4.27%	3.35%	0.99%	-0.06%	0.93%	4.53%	-0.26%	4.27%
-0.3	4.45%	3.48%	1.01%	-0.04%	0.97%	4.61%	-0.16%	4.45%
-0.35	4.55%	3.56%	1.01%	-0.02%	0.99%	4.61%	-0.06%	4.55%
-0.4	4.61%	3.60%	1.00%	0.00%	1.01%	4.57%	0.04%	4.61%

<sup>17</sup>. Although the magnitude of my surplus estimates are much smaller than Borenstein's and much larger than Holland and Mansur's, the ratio of surplus increase between TOU and RTP are remarkably close given the different definitions of TOU used in each case. Borenstein predicted that TOU would have 8-25% the effect of RTP on surplus; Holland and Mansur predicted 15% [7, 10].

<sup>18</sup>. The reason for the lack of monotonicity in consumer surplus can be understood by seeing what happens to the area representing  $\Delta CS$  in Figure 1 with extremely steep, moderate, and extremely flat demand curves. A similar figure should be drawn and examined for the case in which load and price increase with RTP.

Before looking at these results, a regulator might be concerned about charging RTP for customers who have no ability to respond. It would seem unfair to charge customers high RTPs if they could not react. These results indicate that even if customers had no means of knowing or responding to the RTP, the adverse effect of extremely high prices would not cause any problems on average over the year. Flat and countercyclical customers would benefit by not having to subsidize the excesses of others. Even customers with high coincident peak load would not have a large change in average price and could actually save money from other customers' responses. These results indicate that regulators need not worry about the effect of RTP on poor or unresponsive consumers since they will be better off under RTP even if they did not respond.

## 8 Conclusions and Recommendations

The traditional assumption that end users cannot vary their consumption as prices change has led to large, unnecessary investments in peaking plants. In 2006, 15% of the generation capacity in PJM territory ran less than 1.1% of the time (96 hours or less), and 20% of capacity ran less than 2.3% of the time (202 hours or less) [1]<sup>19</sup>. These under-utilized peak generation investments are a luxury that neither providers nor customers want to pay for.

The good news is that the peak load problem can be mitigated by moving flat rate customers onto RTP tariffs. Even with little price responsiveness, surprisingly large peak load reductions can be achieved; at elasticities -0.1 and -0.2, 10.4% and 15.1% respectively can be shaved off of coincident peak consumption. Most other quantities of interest such as generator profitability, overall consumption, and average end user expense will not be affected greatly by a change toward RTP. However, policy makers will be disappointed with the short-term reduction in overall bills. A move toward RTP should be driven by concerns about peak load and equity among end users.

Under current conditions counter-cyclical end users subsidize the high coincident peak loads of others. When problematic, high-peak customers are confronted with higher bills, they will want to make small but important changes. If a peaky customer does not want to alter her consumption habits, then she will face the full price of her own load profile rather than having it subsidized by the rest of the system. Just as consumers have learned to respond to the volatile prices of gasoline, fruits, vegetables, and other commodities, so they can learn to respond to electricity prices. The largest difference is that customers purchase electricity every hour of the year and therefore some customers will want automated devices to react to changing prices.

Because only modest aggregate price elasticities are necessary for large peak capacity savings, most of the benefits can be achieved by shifting only large, responsive customers to RTP. Further, 50% of all possible customer expense savings from load shifting could

---

<sup>19</sup>. This is based on the entire PJM hourly load profile in 2006 [1]. Even at peak load, the system had 17.5% excess available generation capacity. I do not include generation excess at coincident peak load in this calculation because some generation excess is necessary for reliability purposes.

be achieved by shifting only 1.7% of all MWh to another time of day. Large, responsive users are the customers who would benefit the most by installing the equipment necessary for automated response to RTP. With RTP, each customer is free to react in the ways that best serve her interest.

## 9 Acknowledgements

I would like to thank Lester Lave for advising me in this doctoral research. This research was supported in part by the Alfred P. Sloan Foundation and the Electric Power Research Institute through grants to the Carnegie Mellon Electricity Industry Center. My graduate studies are further supported by the National Science Foundation Graduate Research Fellowship Program and the Achievement Rewards for College Scientists Foundation of Pittsburgh.

## 10 References

- [1] PJM Market Data. Available: <http://www.pjm.com/markets/market-monitor/data.html>
- [2] *Assessment of PJM Load Response Programs*. PJM Market Monitoring Unit. Report to the Federal Energy Regulatory Commission, Docket No. ER02-1326-006. August 29,2006. Available: <http://www.pjm.com/markets/market-monitor/downloads/mmu-reports/dsr-report-2005-august-29-%202006.pdf>
- [3] *2005 Price Responsive Load Survey Results*. Available: <http://www.pjm.com/committees/working-groups/dsrwg/downloads/20060615-05-price-responsive-load-survey.pdf>
- [4] Barbose, Galen, Charles Goldman, and Bernie Neenan. *A Survey of Utility Experience with Real Time Pricing*. Lawrence Berkeley National Lab. December 2004. Available: <http://www.osti.gov/bridge/servlets/purl/836966-SZe2FO/native/836966.pdf>
- [5] *Assessment of Demand Response and Advanced Metering*. Federal Energy Regulatory Commission, Staff Report. Docket Number: AD-06-2-00. August 2006. Available: <http://www.ferc.gov/legal/staff-reports/demand-response.pdf>
- [6] Heffner, Grayson C. "Configuring Load as a Resource for Competitive Electricity Markets - Review of Demand Response Programs in the US and Around the World." Lawrence Berkeley National Laboratory. November 2002. Available: <http://repositories.cdlib.org/cgi/viewcontent.cgi?article=1747&context=lbnl>
- [7] Borenstein, Severin. "The Long-Run Efficiency of Real-Time Electricity Pricing." *The Energy Journal*. 2005; 26; 3. pp 93. Available: <http://www.ucei.berkeley.edu/PDF/csemwp133r.pdf>
- [8] *FERC Order on Section 206 Investigation into WECC-Wide Price Cap and the CAISO Ancillary Service Capacity Bid Cap*. February 13,2006. Available: <http://www.caiso.com/179e/179ead8256e10.pdf>

- [9] *RTO-ISO Handbook, Revised - February 28, 2006*. Federal Energy Regulatory Commission. Available: <http://www.ferc.gov/industries/electric/indus-act/rto/handbook.asp#skipnavsub>
- [10] Holland, Stephen P, and Erin T Mansur. "The Short-Run Effects of Time-Varying Prices in Competitive Electricity Markets." *The Energy Journal*, Vol. 27, No. 4.
- [11] Holland, Stephen P. and Erin T. Mansur. *Is Real-Time Pricing Green? The Environmental Impacts of Electricity Demand Variance*. Available: <http://repositories.cdlib.org/cgi/viewcontent.cgi?article=1039&context=ucei/csem>
- [12] Wang, Jing, Nuria Encinas Redondo, and Francisco D Galiana. "Demand-Side Reserve Offers in Joint Energy/Reserve Electricity Markets." *IEEE Transactions on Power Systems*, Vol. 18, No. 4, November 2003. Available: <http://ieeexplore.ieee.org/iel5/59/27910/01245550.pdf?arnumber=1245550>
- [13] King, Chris S, and Sanjoy Chatterjee. "Predicting California Demand Response: How do Customers React to Hourly Prices?" *Public Utilities Fortnightly*, July 1, 2003. Available: <http://www.americanenergyinstitutes.org/research/CaDemandResponse.pdf>
- [14] *PJM Manual 06: Financial Transmission Rights, Revision 8*. Effective Date: March 08, 2006. Available: <http://www.pjm.com/contributions/pjm-manuals/pdf/m06v08.pdf>
- [15] Boisvert, Richard N, Peter Cappers, Charles Goldman, Bernie Neenan, and Nicole Hopper (2007). "Customer Response to RTP in Competitive Markets: A Study of Niagara Mohawk's Standard Offer Tariff," *The Energy Journal*, Vol. 28, No. 1.
- [16] Neenan, Bernie, Donna Pratt, Peter Cappers, and Richard Boisvert. *Does Real-Time Pricing Deliver Demand Response? A Case Study of Niagara Mohawk's Large Customer RTP Tariff*. Lawrence Berkeley National Laboratory. August 2004. Available: <http://eetd.lbl.gov/ea/EMS/reports/54974.pdf>
- [17] *Benefits of Demand Response in Electricity Markets and Recommendations for Achieving Them: A Report to the United States Congress Pursuant to Section 1252 of the Energy Policy Act of 2005*. US Department of Energy, February 2006. Available: [http://www.oe.energy.gov/DocumentsandMedia/congress\\_1252d.pdf](http://www.oe.energy.gov/DocumentsandMedia/congress_1252d.pdf)
- [18] *PJM Overview*. Available: <http://www.pjm.com/about/overview.html>