Wealth Transfers Among Large Customers from Implementing Real-Time Retail Electricity Pricing

*Severin Borenstein**

Adoption of real-time electricity pricing—retail prices that vary hourly to reflect changing wholesale prices—removes existing cross-subsidies to those customers that consume disproportionately more when wholesale prices are highest. If their losses are substantial, these customers are likely to oppose RTP initiatives unless there is a supplemental program to offset their loss. Using data on a sample of 1142 large industrial and commercial customers in northern California, I show that RTP adoption would result in significant transfers compared to a flat-rate tariff. When compared to the time-of-use rates (simple peak/offpeak tariffs) that these customers already face, however, the transfers drop by up to 45%; even under the more extreme price volatility scenario that I examine, 90% of customers would see changes of between a 4% bill reduction and an 8% bill increase. Though customer price responsiveness reduces the loss incurred by those with high-cost demand profiles, I also demonstrate that this offsetting effect is unlikely to be large enough for most customers with costly demand patterns to completely offset their lost cross-subsidy. The analysis suggests that adoption of real-time pricing may be difficult without a supplemental program that compensates the customers who are made worse off by the change. I examine possible "two-part RTP" programs, which allow customers to purchase a baseline quantity at regulated TOU rates, and show they can be used to greatly reduce the transfers associated with adoption of RTP.

The Energy Journal, Vol. 28, No. 2.

* Director, University of California Energy Institute (www.ucei.org); E.T. Grether Professor of Business Administration and Public Policy at the Haas School of Business, University of California, Berkeley (faculty.haas.berkeley.edu/borenste) and National Bureau of Economic Research (www.nber.org). Email: borenste@haas.berkeley.edu.

For helpful comments and discussions, I am grateful to Jim Bushnell, Claude Crampes, Stephen Holland, Rob Letzler, Karen Notsund, Catherine Waddams, Frank Wolak, and three anonymous referees and seminar participants in U.C. Berkeley's Department of Agricultural and Resource Economics and at the U.C. Energy Institute Electricity Camp. My thanks to David Hungerford at the California Energy Commission and Susan McNicoll at Pacific Gas & Electric for helping me obtain access to the data used in this paper.

1. INTRODUCTION

With the restructuring of electricity markets, and the disruptions that resulted in some locations, there has been an increasing focus on efficient pricing of electricity. At the retail level, there have been studies and public policy discussions of real-time pricing (RTP)—retail prices that vary hour-to-hour, reflecting wholesale price variation. Among economists and some policy makers there is widespread agreement on the potential benefits of RTP. There is, however, uncertainty about the full economic impact of RTP. Beyond the real economic costs of implementing RTP—such as installing sophisticated meters and adapting to more complex billing—the resulting wealth transfers also create potential political barriers.

Wealth transfers from moving to RTP would occur because current billing practices—a constant price at all times or simple peak/off-peak pricing (known as time-of-use or TOU pricing)—do not cause retail prices to fluctuate as much as they would under RTP. Under current billing approaches, customers who consume disproportionately high quantities when wholesale prices are high are subsidized by those who consume disproportionately low quantities at those times.

In this paper, I investigate the size of the wealth transfers that would occur if electricity systems were to change from billing large commercial and industrial customers under the simple retail pricing structures currently in use to billing them under a real-time pricing structure. I focus in this paper only on wealth transfers. In other work, I and others have estimated the size of potential efficiency gains from RTP.[1] While efficiency gains are clearly important, policy discussions of RTP proposals can be derailed by concerns that some customers would be harmed significantly by ending the cross-subsidy implicit in the current billing structures. My goal in this paper is to characterize the magnitude of the transfers that would occur with implementation of RTP and explore policies that would reduce the resistance to RTP that could result from those transfers.

The setting I study here is one in which RTP becomes mandatory for all large customers, but the analysis also has implications for so-called voluntary RTP programs. In such programs, if the class of non-joiners is required to cover their costs separately from RTP customers, then this analysis suggests the size of the cross-subsidy that would no longer be available to customers with costly consumption profiles when they choose not to join the RTP program. Thus, for either mandatory or voluntary programs, this analysis indicates how much some customers may have to lose from RTP and how hard they might therefore be expected to fight against it.

Using a dataset of real-time consumption patterns of 1142 large customers in northern California, I analyze their retail electricity costs under alternative billing approaches. I find that moving from a flat rate to RTP would indeed cause significant wealth transfers. Decomposing this effect, however, I find that much of the transfers that would result from a switch from a flat-rate tariff to RTP occur

1. See Borenstein (2005a), Borenstein (2005b), Borenstein and Holland (2005), and Holland and Mansur (2005).

even with movement from flat rates to TOU, a change that has already taken place for most large customers in the U.S.

I then investigate how much the potential losers in a switch to RTP may be able to overcome the loss of cross-subsidy by being price responsive. That is, even if a customer has a relatively costly demand profile—consuming larger quantities at times when the wholesale price is high—it might be able to offset that loss through efficiency gains that occur when it sees the actual real-time electricity price and responds. The results of this analysis, however, suggest that the effect of a customer's own price response on its bill are likely to be small compared to the transfer effect unless the customer exhibits quite high price elasticity.

The results suggest that it may be important to mitigate the wealth transfers from RTP adoption in order to build a broad enough political coalition for its adoption. Where RTP has been broadly adopted, it has usually been structured as a "two-part" pricing program, including a "baseline" component that allows an RTP customer to continue purchasing a fixed quantity of power at a regulated rate. I explain how such programs reduce the wealth transfers that would otherwise accompany RTP adoption and by doing so probably contribute to their acceptance. I then examine empirically the effect of a fairly common type of two-part RTP on reducing wealth transfers that real-time pricing could cause.

In the next section, I explain how I calculate retail rates under alternative billing regimes. In section III, I explain the data used for customer demands and wholesale prices. The results with non-price-responsive RTP customers are presented and analyzed in section IV. In section V, I extend the analysis to allow for price responsiveness by the customers who are charged real-time prices. I discuss the political economy implications of the results in section VI and examine how two-part RTP programs that include a baseline quantity purchased at regulated rates reduce the wealth transfer from RTP adoption. I conclude in section VII.

2. ALTERNATIVE RETAIL BILLING ARRANGEMENTS

Historically, electricity customers have been billed according to one of two general rate designs: a time-independent "flat" electricity price or a time-of-use (TOU) price structure that charges higher rates during pre-designated "peak" times and lower rates at other times. Nearly all large electricity users are charged for energy according to a TOU rate.[2]

Flat electricity rates impose a standard per-kilowatt-hour rate that is charged at all times of the day, week, and year, while it is in force. Time-of-use rates can be as simple as charging a different rate during high demand months than during low-demand months, but in practice are generally more complex. The time-of-use rate structures faced by most of the customers I study here have five different price

2. Customer bills also include a component for transmission and distribution of electricity as well as the energy component. I do not consider the T&D component of the tariff, which often includes a demand charge that is a function of the customer's peak usage during a period, as there is currently little or no policy discussion of changing that.

levels based on the time of year and the day/time of the week. There are two periods during the "winter" months, in effect November through April: the peak rate is in effect from 8:30am to 9:30pm on non-holiday weekdays and the off-peak rate is in effect at all other times. Summer rates, which cover May through October, have three components: Peak period is noon-6pm on non-holiday weekdays; Shoulder period is 8:30am-noon and 6pm-9:30pm on non-holiday weekdays, and off-peak is all other times. Because I have consumption data for hour-long periods, I assume that rate changes occur at 8am and 10pm, rather than 8:30am and 9:30pm.[3]

Clearly, because the wholesale cost of electricity varies by the hour, there are likely to be significant cross-subsidies present in retail rates that vary substantially less often. In addition, there are often cross-subsidies across classes of customers: residential versus commercial/industrial, high- versus low-cost locations, and low-usage versus higher-usage customers. Charging retail customers real-time wholesale energy prices for the electrical energy they consume would eliminate the time-based cross-subsidies within the energy portion of their bills.

In order to focus on the effect of the rate design, I abstract from other subsidies that the political rate-making process might include in the rates. For the group of customers I observe, I assume that each of the rate structures considered—flat, TOU and RTP—raises total revenue from these customers that exactly equals the total wholesale cost of the power they consume.[4] For a given set of wholesale prices, that is sufficient to fully specify the flat and RTP rates.

The TOU rate, however, requires further specification, as there are many different rates for the five TOU periods that would attain the revenue neutrality target. I start by assuming that there is no cross-subsidy across the TOU periods. The resulting TOU rate schedule has a somewhat larger peak-to-offpeak price variation than exists in actual tariffs, however, so I also consider a TOU tariff with a ratio of prices between periods that mirrors the ratios in tariffs that are actually in use.

By comparing the various retail tariffs under the assumption that each generates adequate revenues, I am ignoring any risk premium that might be part of the retail price. It is not clear why such a risk premium would be (or is in fact) built into the electricity tariffs of utilities, because they generally face little revenue-adequacy risk of not be able to recoup their procurement costs from volatile electricity prices. In any case, the risk premium would just shift the net gain of *all* customers by a given amount per MWh.[5]

3. Under this assumption, in the four year period I study, the number of hours each rate is in effect are: winter off-peak, 10,512 hours; winter peak, 6,888 hours; summer off-peak, 10,440 hours; summer shoulder, 4128 hours; summer peak, 3096 hours.

4. In reality, over any given time period revenues from pre-set rates for a group of customers are unlikely to exactly match the utility's costs of serving those customers. Over a four-year period, however, the rate adjustment process is likely to be sufficiently flexible that revenues will deviate very little (on a proportional basis) from costs. Furthermore, there's no obvious reason to believe that this would affect the analysis of transfers among customers other than moving the mean change in bills by the revenue shortfall or surplus.

5 There is the associated cost to customers from bearing this risk, but I have shown in Borenstein (2007) that this risk is easily hedged.

3. DEMAND AND PRICE DATA

In order to estimate the potential size of wealth transfers under RTP, one needs to have data on the demand patterns of individual customers and to analyze how a customer's demand co-varies with the wholesale price of electricity.[6] I have obtained hourly customer-level consumption for 1142 large industrial and commercial customers of Pacific Gas & Electric from January 2000 through December 2003.[7]

I start by using the simplest approach to analyzing transfers, assuming that each customer's demand is completely price inelastic and looking at their payments under alternative billing regimes. I do this for concreteness, as elasticity estimates are controversial and there is no credible way to infer customer-level demand elasticities from the available data. Still, in section V, I assume various levels of demand elasticity—though still the same level for all observed customers—and examine the extent to which introducing such elasticity reduces the losses incurred by those customers that would be harmed by a switch to RTP.

The value of this whole exercise depends on the plausibility of the distribution of wholesale prices assumed. One could use the actual California prices from the same time period as the customer-level data. While these prices have some credibility, there is a real issue of how representative they are of likely prices in the future. In particular the 2000 to 2003 period includes both the California electricity crisis—which ran roughly from June 2000 through May 2001—and the subsequent over-capacity and prices that were widely viewed as having been below long-run equilibrium levels.[8] The 2000-01 crisis resulted in prices that were higher, and possibly more volatile, than normal levels, while the subsequent glut of capacity almost certainly damped peak prices more than off-peak prices, leading to reduced price volatility that would tend to understate the wealth transfer effect that introduction of RTP would have.

While I carry out the analysis using these actual spot prices, I view this as such an unusual period that simulated long-run equilibrium price variation is likely to be more representative of the volatility one should expect in the future. I also study potential transfers that would result using simulated long-run equilibrium wholesale prices. The simulations are based on the model presented in Borenstein (2005a). The model establishes a long-run perfectly competitive equilibrium in capacity and wholesale prices for a given demand profile (load duration curve based on the California ISO control area), assumed aggregate demand elasticity, and

6. In discussing this covariation, I am not suggesting causality, since the customers don't actually face these prices.

7. These are the customers for which PG&E has complete hourly data over this time period. These are among PG&E's largest customers with average annual consumption of 6802 MWh.

8. The emergency building of capacity in response to the California electricity crisis brought online so many new power plants that operators argued prices were then too low to justify further building.

costs of different types of production capacity.[9] The simulation model assures that generators cover their variable plus amortized fixed costs during the sample period. The simulation is of an energy-only revenue model; there are no separate capacity payments. Thus, fixed cost recovery occurs during the highest-demand hours when price exceeds the variable cost of even the least efficient generation units.

The data used for generating the wholesale price series for this paper are not exactly the same as in Borenstein (2005a). First, I use different cost data than those in the earlier paper, reflecting changes in capital and fuel costs since that paper was written.[10] Second, I use only demand data from the 4-year period 2000-2003. By limiting the time period of simulation to just January 2000 through December 2003, I can impose that the resulting prices are sufficient in aggregate to cover the amortized capital and variable costs of all generators during the sample time period.

Absent large elasticity for aggregate demand, much of the capital costs are recovered in peak hours, though exactly how many hours and how peaky the prices are depends on the exact elasticity of aggregate demand. I create two wholesale price series with differing elasticities of aggregate demand and different resulting peakiness of prices. The two simulated scenarios differ in the degree of demand elasticity that within-market producers are assumed to face. Demand elasticity may come about from actual end-user adjustments, but it can also come from import supply elasticity or the system operator utilizing out-of-market resources to provide supply if the market prices rises high enough. With extremely inelastic demand, the simulated market equilibrium includes a very small number of hours in which prices are extremely high. These hours produce the net revenues (scarcity rents) necessary for peaker generation units to cover their amortized fixed costs. With somewhat greater demand elasticity, the long-run equilibrium involves the peaker generators collecting scarcity rents over more hours, but a lower level of scarcity rents and a lower wholesale price in any one of those hours. In scenario I, I assume that the demand elasticity faced by within-market producers is -0.025. In scenario II, I assume an elasticity of -0.1. Summary statistics for the three wholesale price scenarios are presented in Table 1. I focus primarily on the analysis of results from scenario I simulated prices, because price volatility from RTP is greater than under scenario II and is likely to cause the greater wealth

9. All firms are small price takers and all firms earn zero economic profits. There are three different forms of generation capacity: a baseload capacity with high capital cost and low marginal cost; a peaker capacity with low capital cost and high marginal cost; and a mid-merit capacity with moderate capital and marginal cost. Entry/exit occur within each type of capacity until, given the existing capacity and the resulting wholesale prices over time, all firms are breaking even. I show in Borenstein (2005a) that this determines a unique equilibrium.

10. The assumptions I use here for annual production cost are: Baseload (coal) Cost=$208247/MW + $25/MWh; Mid-merit (CCGT) Cost=$93549/MW +$50/MWh; and Peaker (Combustion Turbine) Cost=$72207/MW + $75/MWh. These figures are taken from the PJM (2005), pages 82-83. California does not have coal plants, but (a) there are coal plants in the western grid and (b) the results are not affected substantially by fixing the level of baseload capacity in advance to reflect nuclear and other must-take capacity.

Table 1. Wholesale Prices in Alternative Scenarios (all prices in $/MWh)

Time Period: 2000-2003 (35064 total hours)

		Scenario I Very Volatile Simulated Prices	Scenario II Less Volatile Simulated Prices	Scenario III Actual No. Cal. Spot Prices
Flat-Rate Tariff		93.50	93.41	103.54
Fixed-Ratio TOU Tariff -- Maintaining Actual Price Ratios Among TOU Periods				
Winter	Off-Peak	79.45	79.38	87.98
Winter	Peak	98.42	98.32	108.98
Summer	Off-Peak	79.65	79.56	88.19
Summer	Shoulder	92.96	92.87	102.94
Summer	Peak	151.71	151.56	168.00
Cost-Based TOU Tariff -- Breakeven within Each TOU Period				
Winter	Off-Peak	68.80	68.80	101.16
Winter	Peak	88.57	88.56	110.21
Summer	Off-Peak	74.41	74.90	91.82
Summer	Shoulder	97.07	104.38	107.58
Summer	Peak	203.52	192.50	120.31
Real-time Pricing Tariff				
Minimum Price		65.00	65.00	-285.61
Median Price		90.00	89.13	76.59
Mean Price		88.77	88.77	100.95
Maximum Price		6,321.66	1,051.08	790.00
No. of Hours Price>200		287	918	2,725
No. of Hours Price is Above Highest Simulated Generation Marginal Cost		383	1,713	N/A

transfers. For completeness, I also present results from scenario III, using actual California ISO prices for NP15 during the sample time period, but I do not focus on these because of the highly unusual circumstances surrounding the California electricity crisis.[11]

For each of these price series, I also had to create a flat-rate tariff and time-of-use tariffs as the comparison points for calculating the transfers. To do so, I considered the 1142 customers as a distinct customer class and calculated the rates, flat and TOU, that would exactly cover the wholesale cost of acquiring power for this customer class. For each wholesale price series, I have calculated a single break-even flat rate. I have calculated two different sets of TOU rates. The first permits no cross-subsidy across the five TOU periods; I refer to this as "Cost-Based TOU" or TOU-C. The second places a constraint on how much TOU rates can vary between periods so the ratio of rates between periods is approximately

11. Still, it is worth pointing out that scenarios I and II assume equilibrium zero-profit capacity investment and no market power. In reality, capacity investment will almost always differ from the exact breakeven level—sometimes too high and other times too low—and even with the greater understanding of market power policymakers now have, one certainly cannot be assured that it has been permanently eliminated.

equal to the ratio in PG&E's most common TOU tariffs; I call this "Fixed-Ratio TOU" or TOU-F. All or these tariffs are set to assure that the revenue received exactly covers the wholesale cost of power for this set of customers as a class.

Table 1 presents the flat retail price and TOU retail prices under the three different wholesale price distribution assumptions, as well as data on the distribution of wholesale prices for each case. Scenarios I and II, with simulated prices, reflect different degrees of wholesale price volatility depending on the elasticity of the aggregate demand faced by producers in the market. With less elasticity, *i.e.,* scenario I, peaker generating plants recover their fixed costs in fewer hours with higher prices; the peak price is substantially higher in scenario I than scenario II and the price is above the marginal cost of the simulated peaker generators in substantially fewer hours.

Scenario III uses the actual wholesale prices from the California ISO's real-time balancing market for the area in which the observed customers are located. I do not include information on the number of hours in which prices exceeded the marginal cost of peaker generation, because there is no reason to think that generators actually earned rents that exactly covered their amortized capital costs during this period.

Certainly, the assumptions behind the simulated price series, as well as the unique circumstances that generated the actual wholesale prices, cast doubt on whether any one of these series actually is likely to occur in the future. The issue is not whether these simulations are reflective of the actual market institutions or aggregate demand function or elasticities, but whether the prices reflect wholesale (and RTP) prices that are likely to result. The point is that the three scenarios span a range of price patterns and volatility that would still allow generating companies to fully cover their costs. So, the results give a sense of the range of possible outcomes, as well as the robustness of the general conclusions to a variety of plausible price patterns.

I present these scenarios separately from the later analysis in which the *observed* customers are assumed to be able to demonstrate some price elasticity in order to distinguish between two effects that will mitigate the size of transfers. The first effect is from aggregate demand (or import supply) elasticity that damps price volatility, as is demonstrated in the difference between scenarios I and II, the effect of which is discussed in the next section. The second effect is from a customer itself responding to volatile prices by reducing consumption at peak times and increasing consumption off peak. For the observed customers, I ignore this second effect in section IV, but return to it in section V.

4. TRANSFERS FROM RTP ADOPTION IF CUSTOMERS ARE NOT PRICE RESPONSIVE

I calculate the electricity bills for each of the 1142 customers in the dataset under the four alternative billing arrangements using each of the wholesale price scenarios. The bills include a flat charge for transmission and distribution of $40/MWh. The T&D charge has no effect on the magnitude of the transfers, but I include it in order to give a more accurate picture of the proportional impact from changing billing arrangements on the customer's electricity bill. On average T&D comprises somewhat less than half of the electricity bill, so the proportional changes in just the energy component of customer bills are slightly less than twice as large. Throughout the calculations in this section, I assume that customers make no change in their consumption in response to changes in the billing arrangement.[12]

For each wholesale price scenario, each customer's payments under RTP can be compared to their payments under a flat rate billing arrangement with the same wholesale price series. Of course, some customer bills increase compared to flat rates and others decrease, while the total revenue collected from this class of customers is, by construction, held constant. The distribution across customers of percentage gains and losses is shown in the first line under each scenario in Table 2. The more volatile simulated prices (scenario I) result in transfers under RTP that have more extreme upper and lower distribution tails. Under all three scenarios, there could be substantial winners and losers. Under scenario I, one-quarter of the

Table 2. Distributions of Change in Customer Bills Compared to Flat-Rate Tariff (1142 customers over 4 years)

	PERCENTILES									Absolute Transfer
	1st	5th	10th	25th	50th	75th	90th	95th	99th	Per Customer
Scenario I: Very Volatile Simulated Wholesale Prices										
RTP	-13%	-6%	-4%	-2%	1%	8%	12%	15%	19%	$125,409
TOU-C	-10%	-5%	-4%	-2%	0%	5%	9%	10%	15%	$89,173
TOU-F	-5%	-3%	-2%	-2%	0%	3%	5%	6%	9%	$51,873
Scenario II: Less Volatile Simulated Wholesale Prices										
RTP	-12%	-5%	-4%	-2%	1%	6%	10%	12%	19%	$106,920
TOU-C	-9%	-4%	-3%	-2%	0%	5%	8%	10%	14%	$86,097
TOU-F	-5%	-3%	-2%	-2%	0%	3%	5%	6%	9%	$51,822
Scenario III: Actual Northern California Wholesale Prices										
RTP	-12%	-4%	-3%	-2%	1%	4%	8%	14%	33%	$150,646
TOU-C	-2%	-1%	-1%	-1%	0%	1%	2%	3%	4%	$23,931
TOU-F	-5%	-3%	-2%	-2%	0%	3%	5%	6%	9%	$57,440

Distributions of Change in Customer Bills from TOU-F to RTP (1142 customers over 4 years)

	PERCENTILES									Absolute Transfer
	1st	5th	10th	25th	50th	75th	90th	95th	99th	Per Customer
Scenario I	-9%	-4%	-3%	-1%	1%	4%	7%	8%	15%	$80,397
Scenario II	-8%	-3%	-2%	-1%	0%	3%	5%	6%	14%	$59,203
Scenario III	-15%	-5%	-4%	-1%	0%	2%	6%	13%	32%	$133,847

12. By assuming that T&D is charged at a flat rate, I ignore the impact of demand charges that are part of T&D. Demand charges are a fee calibrated to the customer's peak usage during a given period. To the extent that a customer's usage is correlated with system demand, demand charges increase the size of transfers associated with departing from flat rate electricity billing.

customers see their bills rise by 8% or more (as indicated by the 75th percentile change in bills). Comparing scenario II with scenario I, it is evident that even with less volatile wholesale prices, transfers are nearly as large.

The average absolute size of a customer's gain or loss is indicated by the right-hand column of Table 2. This column shows the average per-customer *absolute value* of bill changes over the 1142 customers. A switch from flat rates to RTP under scenario I, for instance, would bring about an average bill change of $125,409 over the four-year period. The first row of scenario II shows that the average transfer would be $106,920 with that set of wholesale prices, only 15% smaller than in scenario I. The transfers would be larger if the actual wholesale prices that obtained during this period, scenario III, were indicative of future prices.

Even continuing to ignore the price-dampening effect of RTP, however, the actual bill changes would be much smaller than the figures in the right-hand column. The reason is that all of these large customers are already on TOU rates. So, from a political economy viewpoint, the relevant change is from TOU to RTP. As mentioned earlier, I use the five TOU time periods (3 periods in summer, 2 in winter) under two different sets of TOU rates for this "customer class" of 1142 customers. TOU-C rates are set so that each time period meets its own separate revenue requirement; TOU-F rates meet the revenue requirement overall while maintaining preset percentage price differences among the time periods.

Focusing first on scenario I, it is clear that a shift from flat rates to TOU pricing imposes a significant proportion of the transfers that would occur by moving all the way to RTP, but what proportion depends very much on the type of TOU. TOU-C involves larger price differentials between the periods than TOU-F. Starting from flat rates, TOU-C causes most of the transfers that would result from RTP. On average, TOU-C results in a customer gaining or losing $89,173, 71% of the transfers that would result from full RTP. TOU-F prices vary less and, as a result cause smaller transfers compared to flat rates. The $51,873 average transfer under TOU-F in scenario I is 41% of the average level that would result from full RTP.

The bottom panel of table 2 presents the transfers that would result from switching from TOU-F to RTP for these customers. From a political economy viewpoint, this may be the most relevant comparison, because TOU-F most closely reflects the current billing arrangement for these customers. Under scenario I, the aggregate transfers are 36% smaller with this switch than under a flat-rate to RTP switch.

Under wholesale price scenario II, the story is very much the same, except the effect of TOU prices is more like RTP, because RTP prices don't have as extreme price spikes under scenario II as under scenario I. With scenario II, both TOU-C and TOU-F result in transfers that are a higher proportion of RTP transfers (81% and 48%, respectively) than occurs in scenario I. Under scenario II, the aggregate transfers are 45% smaller with a TOU-F to RTP switch than with a flat-rate to RTP switch.

The results of using TOU-F under the scenario III, the actual wholesale prices, is odd because the inter-period price differences maintained under TOU-F are actually larger than the ratio that would result from each TOU period breaking even. TOU-F prices would cause larger transfers than would have occurred under TOU-C. The very large transfers under RTP in scenario III are driven more by inter-year price variation than normal peak/off-peak price variation. The average annual prices per MWh in the four years were $155, $107, $66 and $75, compared to $107, $80, $82, and $86 under simulated scenario I and $98, $82, $85, $90 under simulated scenario II.

In considering the size of transfers from a switch to RTP, one might be especially concerned with outliers, particularly customers that would stand to lose large amounts from a switch to RTP. There seem to be fairly few such cases. Looking at a switch from TOU-F to RTP, under price scenario I, only 11 of the 1142 customers would stand to lose more than $100,000 per year (and none would lose more than $300,000 per year). For scenarios II and III, the number of customers losing more than $100,000 per year is 7 and 27, respectively.

Unfortunately, for confidentiality reasons, I do not have any data on the type or location of these customers, so it is not possible to characterize the type of customers—within this set of large industrial/commercial customers—that are more likely to be harmed. In both of the simulated price scenarios, however, it is worth noting that the winners on average are significantly larger customers than the losers. The average electricity consumption by firms that would see their bills decrease due to a switch from TOU-F to RTP is about twice as great as the average electricity consumption by firms that would see a bill increase. Regressing the percentage bill change from TOU-F to RTP on total electricity consumption yields a statistically significant (at the 1% level) negative estimate. The estimate suggests that a one standard deviation increase in customer size is associated with about a 0.8% smaller bill resulting from a switch to RTP. This isn't particularly surprising, because the heaviest electricity consumers tend to be industrial processing plants that have fairly constant electricity demand and are most likely to benefit from RTP. Still, the low R^2 of these regressions—0.03 under scenario I and 0.02 under scenario II—suggest that there are many other factors that affect cross-sectional variation in costliness of demand patterns.[13]

Finally, the transfers calculated here can be decomposed into within-year and between-year components, where the former is calculated as a customer's gain or loss on RTP compared to a different flat or TOU rate each year that exactly covers the wholesale cost of power in that year for the sample customers. The difference between the customer's overall gain/loss and the within-year component is the between-year contribution. This between-year effect could result because some customers have consumed more electricity, due to increased customer activity or simply bad luck, in a year in which wholesale prices were high (*e.g.,* the year 2000 in scenario III). In predicting transfers and opposition to RTP, this may

13. The same pattern holds under scenario III, the actual northern California prices, but the effect is smaller and significant only at the 11% level.

be of less importance.[14] Focusing only on within-year transfers would change the conclusions relatively little under scenario I and very little under scenario II. A switch from TOU-F to RTP causes the *within-year* absolute transfer per customer under scenario I to be $67,185 compared to the overall absolute transfer per customer of $80,397. Under scenario II, the within-year effect is $54,370 compared to $59,203 overall. Under scenario III, however, the within-year effect is much smaller than the overall transfer, $57,428 compared to $133,847 overall. This reflects the fact that all customers paid much higher rates in 2000 and 2001 than in 2002 and 2003 under scenario III.

5. TRANSFERS FROM RTP ADOPTION IF CUSTOMERS RESPOND TO PRICE VOLATILITY

The calculations in the previous section assumed that the observed customers would not change their consumption in response to changes in retail electricity prices. The results highlight how, apart from any response of the RTP customers, the volatility of wholesale prices will affect the size of the transfers. Of course, the whole point of RTP is for customers to respond and, by doing so, to increase the efficiency of the entire electricity system. In other work, I have examined the systemwide efficiency of such consumption changes. Here, I examine the effect of price response just on the surplus that these customers would receive, and in particular whether the gains from price response would substantially lessen the losses that some customers would otherwise incur with a switch to RTP.[15]

In order to analyze the benefits or losses to customers when they exhibit price elasticity, it is necessary to analyze consumer surplus instead of simply the total payments by customers. Total payments would fail to capture the benefits to consumers when they increase consumption during low-price hours and would misstate the losses when a customer reduces its bill by lowering consumption during high price periods, but also loses the value of that consumption.

The consumption actually observed for these customers occurred when they were facing a billing regime that most closely resembled TOU-F, so I use that as the baseline from which changes in consumer surplus are measured. I then

14. Still, the between-year component would not be zero even if customer consumption patterns didn't vary year to year. If one customer tends to consume more on high demand days, then it is going to be harmed more in years in which system demand is high and this will show up in part as a between-year transfer.

15. For comparability to the results of the previous section, I use the same distribution of wholesale prices as before. Implicitly, I am assuming that adoption of RTP by the customers in this sample would not change the distribution of wholesale prices. I do this in order to maintain the clear distinction between wealth transfer effects caused by the distribution of wholesale prices that is exogenous to any one customer and the mitigation of the effect that is possible through the customer responding to those wholesale prices. In aggregate, the assumption is unlikely to hold; increasing the share of customers on RTP would dampen price volatility. The way to incorporate this effect in analyzing the wealth effect on any one customer moving to RTP would be simply to assume a more elastic aggregate demand, and thus a less volatile wholesale price series, whether or not the observed customer were to switch.

consider possible changes from the observed consumption under alternative assumptions about the customer's price elasticity of demand.

To be concrete, I assume that customer *i* quantity demand in hour *h* has a constant elasticity with respect to price in hour *h* so that $q_{hi}=A_{hi}\,P_h^{\,e}$. I assume a certain elasticity, e, and can then derive A_{hi} for each customer-hour based on the assumption that q_{hi} is its observed consumption and P_h is the TOU-F price for that hour. The customer's change in consumer surplus from facing an RTP price in hour *h* rather than the TOU-F price for that hour would then be:

$$\Delta CS_{hi} = (A_{hi}/(e+1))\,[P^{(e+1)}_{TOU\text{-}F} - P^{(e+1)}_{RTP}] \qquad [1]$$

Aggregating the result of [1] over all hours for a customer allows me to calculate its change in consumer surplus. As a basis for comparison, I then divide the customer's change in consumer surplus by its total bill under TOU-F pricing and observed consumption.

The distributions of the results are presented in Table 3. I show results for only scenarios I and II.[16] The first row in each scenario section indicates the distribution of percentage change with no customer price response. This row just matches the bottom panel of table 2 except with a reversal of the sign because I am now considering the change in consumer surplus rather than expense.[17]

The remaining three rows in each section present the distribution of change in consumer surplus with varying levels of assumed price elasticity of demand on the part of the observed customers. Unlike the results with zero elasticity from these customers, with price responsiveness, the aggregate consumer surplus change over all customers is not zero. By revealed preference, each customer is at least as well off as if it exhibited no price response. Thus, with price response, the *aggregate* change in consumer surplus is positive for this class of customers.

The results presented in table 3 indicate that while price responsiveness will mitigate to some extent the losses of customers with costly demand profiles, it may not substantially change the political economy of the issue. Modest price elasticity does not have as large an effect as one might hope on a customer's net gain from RTP. For instance, under wholesale price scenario I, with no price responsiveness, a customer at the 10th percentile of the distribution sees a consumer surplus loss of 7% of its TOU-F bill. But even if customers have a -0.1 price elasticity in response to RTP price variation, the customer at the 10th percentile still sees a loss of 4% of its TOU-F bill. Looking across table 3, it is clear that an elasticity of -0.1 moves the distribution of gains/losses in the positive direction, but by only a few percentage points or less.

16. Under scenario III, as suggested earlier, the transfers due to RTP are driven in large part by the inter-year price variation due to California electricity crisis. The negative spot prices that actually occurred make it impossible to calculate equation [1] for scenario III, but the basic result of this analysis would certainly carry over.

17. In these cases the dollar value of the change in consumer surplus is the same as in the bottom panel of table 2. In the cases with price response, the dollar value of changes in consumer surplus are proportionally smaller.

Table 3. Distributions of Change in Customer Consumer Surplus as a Result of Switching from TOU-F to RTP Tariff (1142 customers over 4 years)

Assumed Customer Elasticity	PERCENTILES 1st	5th	10th	25th	50th	75th	90th	95th	99th	Share of Customers with ΔCS>0
Scenario I: Very Volatile Simulated Wholesale Prices										
e =0	-15%	-8%	-7%	-4%	-1%	1%	3%	4%	9%	41%
e =-0.025	-14%	-7%	-6%	-4%	0%	1%	3%	4%	9%	47%
e =-0.1	-12%	-5%	-4%	-2%	1%	3%	4%	5%	10%	62%
e =-0.3	-6%	0%	1%	2%	4%	6%	7%	7%	10%	95%
Scenario II: Less Volatile Simulated Wholesale Prices										
e=0	-14%	-6%	-5%	-3%	0%	1%	2%	3%	8%	45%
e=-0.025	-14%	-6%	-5%	-3%	0%	1%	2%	3%	8%	48%
e=-0.1	-13%	-5%	-4%	-2%	1%	2%	3%	3%	9%	56%
e=-0.3	-10%	-3%	-2%	0%	2%	3%	4%	5%	9%	77%

The right-hand column of table 3 indicates the share of customers who benefit from a switch to RTP under the assumed level of customer price elasticity, and continuing to take the wholesale price distribution as exogenous. With no customer price elasticity, about 41% of customers would gain consumer surplus as a result of switching from a TOU-F retail tariff to RTP. Demand elasticity increases this number and shifts the distribution, but even if these customers have an elasticity of -0.1, 38% of customers are still worse off.[18] Only with a much higher elasticity, which seems unlikely in the short run, are a large majority of these customers likely to benefit from a switch to RTP.

Demand elasticity has a greater effect on customer gains in scenario I, in which wholesale prices are very spiky, than under scenario II, which has more moderate spikes. This reflects the fact that the surplus gain from elasticity is larger when prices are more volatile.[19] Unfortunately, many customers still have a very negative view of price volatility, seeing it as introducing detrimental risk into the firm's operations. In other research with these same data, I have investigated the potential for mitigating that risk using straightforward hedging instruments.[20]

6. TWO-PART RTP PROGRAMS LESSEN THE WEALTH TRANSFERS FROM RTP ADOPTION

The analysis suggests that adoption of real-time pricing may be difficult without some supplemental program that compensates the customers who are made worse off by the change. Georgia Power, which runs the oldest and largest RTP program in the U.S., has mitigated the lost cross-subsidy effect by allowing

18. The magnitudes of these aggregate consumer gains are consistent with the effects that Borenstein (2005a) and Holland and Mansur (2006) find.

19. Borenstein and Holland (2005) make this point in evaluating the gains for an individual customer that moves to RTP.

20. See Borenstein (2007).

customers to lock in a certain baseline level of consumption at the regulated TOU rate and pay RTP only for deviations from their baseline level of consumption. Such "two-part" RTP pricing programs are often touted for their risk mitigation effect, but they can also maintain the cross-subsidy and thus potentially reduce political opposition to RTP.[21]

The way in which such two-part RTP programs allow maintenance of the pre-RTP cross-subsidy may not be obvious at first. In a TOU program, if the retail price of power during each TOU period is set equal to the true cost for power during that period, then how could a customer be cross-subsidized by being allowed to purchase at that price? The answer is two-fold.

First, TOU prices frequently do not actually reflect the true peak/off-peak difference in wholesale costs, instead underpricing the peak period and overpricing the off-peak period. In such case, assigning a customer baseline (CBL) level for TOU-rate purchases based on the customer's past levels of consumption during each TOU period maintains the average cross-subsidy that the customer received under the pre-RTP plan due to the cross-subsidy *between* TOU periods. Those with disproportionate consumption during the designated peak period will continue to benefit from the fact that they consume disproportionately at times when the retail price of the energy is on average below the wholesale cost.

A second, closely related effect is somewhat more subtle: there is a *within* TOU period cross-subsidy that is usually maintained. If under two-part RTP a customer is permitted to buy a baseline demand pattern within a TOU period that is more costly than the retail provider's average acquisition cost for power it buys during that TOU period as a whole, then the customer will continue to be cross subsidized. For example, consider a summer peak TOU period that covers noon-6pm for all non-holiday weekdays during May-October. Assume that the TOU price is set by the retail provider to cover the expected wholesale cost of the power acquisition it needs to make during that period, so there is no cross subsidy between TOU periods. Consider a customer that has disproportionately high demand (compared to the retail provider's load) during August peak periods, which happen to be when wholesale prices are highest in the summer. Under a TOU program, that customer is cross-subsidized because *within* the summer peak TOU period it is buying a disproportionate quantity of power during the most expensive wholesale price hours.

This within-TOU-period subsidy could continue or be eliminated under a two-part RTP program depending on the way in which the CBL quantity—the quantity the customer is allowed to purchase at the TOU rate—is determined. If the CBL quantity is proportional in each hour to the retail provider's aggregate load, then all customers are buying the same standardized product (though differing amounts of it) at a cost-based price and there is no within-period cross-subsidy. On the other hand, if each customer is allowed to customize its baseline quantity purchased across the hours that are within each TOU period—the most

21. See Barbose, Goldman and Neenan (2004) for a broad survey of RTP programs with alternative baseline approaches.

common system being a baseline determined by the customer's own consumption in past years on the equivalent dates—then some customers will purchase more quantity during the most expensive hours. If those customers are still allowed to buy that quantity at the standard TOU rate—as is the case in most programs—then the cross-subsidy will be maintained. In practice, nearly all CBLs have customized the demand pattern of each customer's baseline consumption to reflect the customer's past consumption, thus maintaining a significant cross-subsidy.

Two-part RTP programs can in fact be designed to eliminate all cross-subsidy in energy costs. This can be done while still allowing customers to pre-purchase fixed quantities of power in order to reducing the risk of volatile power bills, an effect that is completely distinct from the impact on transfers. It is simple to describe two such proposals that happen to represent opposite extremes of flexibility.

The first, as suggested above, would offer a standard product within each TOU period that is a fixed-proportion in each hour of the retail provider's expected aggregate load. The retail provider would price this at the expected cost of the "load slice." No "cherry picking" of the most expensive hours would be possible, because the only product available would be a fixed proportion of aggregate demand in all hours. So, no cross-subsidy would result.

An alternative proposal offers complete flexibility, but sets separate prices for each hour. Under this proposal, the retail provider would set a forward price for each hour of the coming year based on its best forecast of wholesale price (through an analysis of expected demand and supply drivers). The retail provider would price each hour so that there is no expected cross-subsidy across hours. A customer would then be allowed to craft its own baseline quantity, potentially buying forward a different quantity for each hour of the year.[22]

Each of these two proposals permits the risk management function that is often suggested as the basis for two-part RTP programs while eliminating the cross-subsidies the exist under the pre-RTP system. Yet, neither of these programs has been adopted anywhere in the U.S. Instead, customized baselines have been used or baselines purchases have simply been made available at prices below the expected wholesale costs. This suggests that the designs of the customer baseline programs have been aimed at mitigating the loss of cross-subsidy as well as reducing the perceived bill risk associated with RTP.

Finally, note that even the most sophisticated two-part RTP program is unlikely to completely eliminate wealth transfers that would result from RTP adoption. Even if every customer were required to pre-purchase their expected hour-by-hour demand at the regulated TOU rates, the stochastic components in consumption and real-time wholesale prices would cause some transfers. On the unexpectedly hot summer day, prices would rise and those customers who see the greatest increase in their consumption above their expected level would be hurt the most. Thus, those customers whose demand experiences unanticipated shocks that are most strongly positively correlated with shocks to system demand and price would still take a wealth hit. This would be the case even under a two-part

22. Borenstein (2005b) describes this proposal in greater detail.

RTP program in which each customer buys in advance their customized expected demand quantity for each hour.

The difficulty of eliminating transfers with a two-part RTP program is illustrated by a simple calculation of a representative two-part RTP programs. I present here results from an analysis of a hypothetical two-part RTP program for the same 1142 customers.

I consider the effect of a two-part RTP program in which every customer purchases its expected demand for every hour of the four-year sample at the TOU-F price. This generates a capital gain or loss relative to the real-time price that obtains in each hour and that gain or loss is added to—in most cases partially offsets—the gain or loss from the switch to RTP. This is financially equivalent to a customer pre-purchasing its expected demand at the TOU-F price and then being charged or refunded for deviations from its expected demand at the real-time price. Throughout this analysis, I again assume that these customers exhibit zero price elasticity.

The customer's expected demand during a given hour for this analysis is taken to be its average demand in that month-weekday/weekend-hour of day over the four-year sample. That is, every month *e.g.,* January, covering the four January's in the sample) generates 48 hourly expected demands, 24 for the 24 hours of weekdays in the month and 24 for the 24 hours of weekends (and holidays) in the month. So, for instance, a customer would pre-purchase at TOU-F prices its average demand in 3-4pm January non-holiday weekdays for each of the 3-4pm January non-holiday weekdays that appears in the sample (about 80 such hours in the sample).

The results of this analysis are shown in Table 4. For both simulated price scenarios, table 4 reproduces the distribution of transfers for a switch from TOU-F to RTP that was shown in Table 2. Then it presents the transfers that result when the RTP is paired with pre-purchase of expected demand at TOU-F prices, the two-part program. The two-part program reduces the transfers by roughly half, but that still leaves substantial gains and losses. In the case of the more volatile price scenario, adding the two-part component reduces transfers by about 40% (from an average gain or loss of $80,397 to an average of $48,495). It is more effective in scenario II, because lower price volatility means there is less unpredictable price/demand correlation that the pre-purchase cannot cover,

Table 4. Distributions of Change in Customer Bills from TOU-F to RTP or 2-Part RTP (1142 customers over 4 years)

	PERCENTILES									Absolute Transfer
	1st	5th	10th	25th	50th	75th	90th	95th	99th	Per Customer
Scenario I - RTP	-9%	-4%	-3%	-1%	1%	4%	7%	8%	15%	$80,397
Scenario I - 2-part	-5%	-2%	0%	0%	1%	2%	4%	5%	12%	$48,495
Scenario II - RTP	-8%	-3%	-2%	-1%	0%	3%	5%	6%	14%	$59,203
Scenario II - 2-part	-2%	0%	0%	0%	1%	1%	2%	3%	6%	$25,162

as discussed above. In that case, the two-part component eliminates about 58% of the transfers.

7. CONCLUSIONS

Introducing real-time electricity pricing is likely to harm some customers by removing the existing cross-subsidies to customers that consume disproportionately more when wholesale prices are highest. Those customers are likely to oppose RTP initiatives if their potential loss is substantial and there is no supplemental program to offset their loss. Using data on a sample of 1142 large industrial and commercial customers in northern California, I've shown that implementing RTP results in significant transfers compared to a flat-rate tariff. Forty to seventy percent of these transfers, however, occur with just a change from flat-rate to time-of-use pricing, a change that has already taken place for the customers in this sample, and for most large industrial and commercial customers in the U.S. Still, current TOU tariffs probably understate the long-run equilibrium cost differential between peak and off-peak periods, thus reducing the transfer caused by such rates and increasing the additional transfer that would result from moving to full RTP.

One hope for broader RTP support is that customers may help themselves under RTP by reducing consumption when prices are high and consuming more when prices are low. While this price responsiveness generates substantial efficiencies in aggregate, I demonstrate that it is unlikely to be large enough for most customers with costly demand patterns to overcome their lost cross-subsidy. Even if customers exhibit real-time price elasticities of -0.1, I conclude that a large share of them would still be losers under RTP.

The analysis makes clear that in the political economy of retail electricity pricing there is likely to be a role for programs that mitigate the wealth transfers from RTP adoption while still achieving the efficiency gains. I've shown that "two-part" RTP programs, which allow customers to buy a baseline quantity at a regulated rate, can fulfill this function under their typical implementation.

REFERENCES

Barbose, Galen, Charles Goldman and Bernie Neenan (2004). "A Survey of Utility Experience with Real-Time Pricing," Lawrence Berkeley National Laboratory Working Paper No. LBNL-54238, December.

Borenstein, Severin (2005a). "The Long-Run Efficiency of Real-Time Electricity Pricing," *Energy Journal,* 26(3): 93-116.

Borenstein, Severin (2005b). "Time-Varying Retail Electricity Prices: Theory and Practice," in Griffin and Puller, eds., *Electricity Deregulation: Choices and Challenges*, Chicago: University of Chicago Press.

Borenstein, Severin (2007). "Customer Risk from Real-Time Retail Electricity Pricing: Bill Volatility and Hedgability," *Energy Journal,* 28(2): 111-130

Borenstein, Severin and Stephen P. Holland (2005). "On the Efficiency of Competitive Electricity Markets With Time-Invariant Retail Prices," *RAND Journal of Economics*, 36(3):469-493.

Holland, Stephen P. and Erin T. Mansur (2006). "The Short Run Effects of Time-Varying Prices in Competitive Electricity Markets" *Energy Journal,* 27(4): 127-155.

PJM Interconnection Market Monitoring Unit (2005). "2004 State of the Market Report." Available at http://www.pjm.com/markets/market-monitor/downloads/mmu-reports/pjm-som-2004.pdf.