FSC

FREEMAN, SULLIVAN & CO.

A MEMBER OF THE FSC GROUP

2012 San Diego Gas & Electric Peak Time Rebate Baseline Evaluation

Prepared for:
San Diego Gas & Electric
Leslie Willoughby
Kathryn Smith

April 2013

Prepared by:
Stephen George
Josh Bode
Dries Berghman

Freeman, Sullivan & Co.
101 Montgomery St., 15th Fl
San Francisco, CA 94104
fscgroup.com

The FSC Group

Table of Contents

1 Executive Summary

In 2012, San Diego Gas & Electric Company (SDG&E) enrolled approximately 1.2 million residential premises in a peak time rebate (PTR) program, branded as "Reduce Your Use Rewards." Along with Southern California Edison's (SCE) default PTR program, which has been implemented for all customers with smart meters (about half of SCE's residential base), SDG&E's program is one of the two largest implementations of default PTR in the United States.

PTR is a "pay for performance" program that pays consumers to reduce electricity use during the peak period on selected days (referred to as event days) that are not known until the day before they occur. The incentive is paid based on the difference between metered load during the peak period on event days and an estimate of what the customer would have used during the same period if the PTR event had not occurred. This estimate is referred to as baseline load. The accuracy and magnitude of incentive payments is dependent on the accuracy of the baseline estimate. Given the normal fluctuation in usage across days, it is very difficult to accurately estimate baselines for individual customers on individual event days.

This report presents the results of an analysis of the accuracy of baseline estimation methods for SDG&E's residential PTR program. It presents baseline accuracy results for 21 different baseline models, with and without 10 different approaches to same-day adjustment and for 2 event windows, for a total of 420 different baseline/event-window combinations. The baseline types studied in this report include the method currently used by SDG&E's program, which estimates peak period load be averaging the load for each customer across the highest three out of the prior five non-event weekdays for weekday events and three of the highest five weekends for a weekend event. This is referred to as a 3/5 baseline method. Numerous other day-matching and weather-matching baselines and regression analysis were also examined.

In addition to analyzing baseline accuracy, this report also examines the accuracy of load impacts and payments made to customers based on the estimated load reductions using each baseline method. Baseline, impact and payment accuracy are related but are not the same thing. In percentage terms, impact errors are always larger than baseline errors; when the event impact is relatively small, the impact error tends be much larger than the baseline error. Furthermore, because customers are paid for estimated load reductions but are not penalized for estimated load increases, baseline errors that lead to overpayments are not always offset by underpayments. This asymmetry in payment error is larger when average load reductions are small, as they are with SDG&E's Reduce Your Use program.

Key findings concerning baseline accuracy include:

- While some methods produce reasonably accurate baseline estimates across all customers and all event days, no method is close to being accurate for individual customers on individual event days. Even when averaging across multiple event days, all methods tend to systematically over or underestimate reference load for a substantial share of customers. For example, the best performing baseline, a weather-matching baseline using five days with similar weather, generates baseline errors greater than 24% or smaller than -24% for 20% of customers on the average event day (calculated using 15 event days).
- Individual customer regressions and weather matching algorithms (with adjustments) produce more accurate baseline estimates than any other method; however, they are still highly inaccurate for individual customers and individual events.
- While the regression method and a few weather-matching baseline methods provide marginal improvements in baseline accuracy, the current 3/5 baseline performs relatively well. Out of the 21 baseline types that were tested, the 3/5 baseline generally ranks in the top quarter in

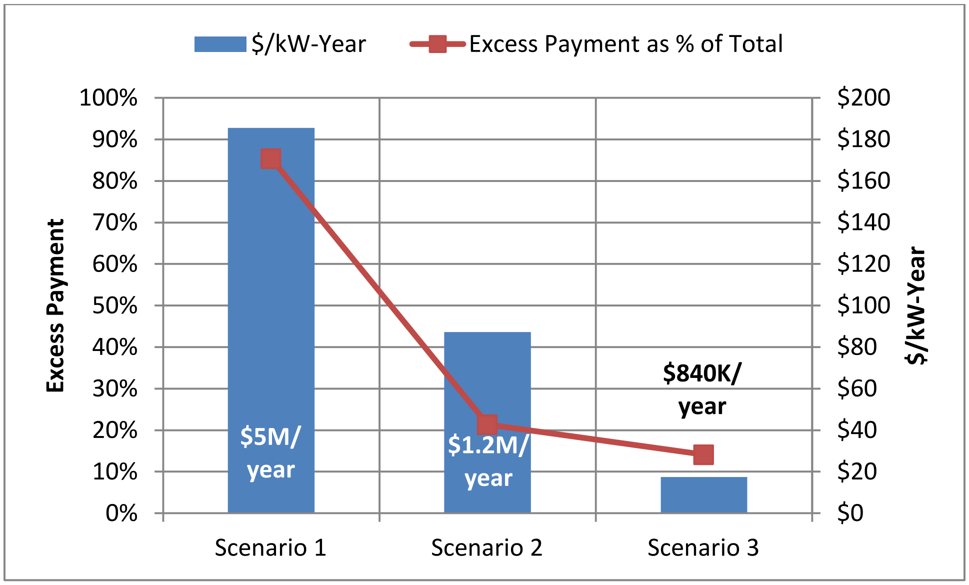
all baseline accuracy metrics. For example, for 80% of customers, the 3/5 baseline has baseline error ranging from -31% to 28% on the average event day, whereas the best baselines manage to achieve errors ranging from -24% to 24% on the average event day for these same customers. Thus, the best baseline only represents a small improvement over the current 3/5 baseline.

- While small same-day adjustments can provide additional improvements in baseline accuracy, these improvements may not hold in a real-world situation in which customers might pre-cool their homes in anticipation of a PTR event or reduce their load in anticipation of the event. Consequently, we do not recommend adopting a same-day adjustment. The marginal gain in accuracy based on simulation analysis could easily be offset by errors caused by logical changes in consumer behavior if the assumptions underlying the simulation (that customers do not change usage before the event) are wrong.
- Gains in accuracy from switching away from the current 3/5 baseline method are small and we do not recommend that SDG&E use a different baseline method.

Key conclusions concerning payment accuracy and program design include:

- Baseline and payment errors result in payments being made to many customers who do not reduce demand. These payment errors must be recovered from all customers.
- Payment error is largest when demand reductions per customer are small, as they are for SDG&E's default program. When assuming an average load reduction of 2% for all customers, payment error accounts for 93% of all payments made when using the current baseline. In other words, if PTR settlement were completely accurate, rebate costs would be 93% lower than they would be using the current, "relatively accurate" baseline. However, when average impacts are larger, say 10% for example, the percent of total payments made in error drops dramatically – in this case, to 44%. With larger impacts, over and under payments partially offset each other across customers and event days, but not enough to generate the correct average payment.
- Changes in payment rules can reduce costs and/or excess payments for any baseline method. Numerous payment rule changes were examined, including setting thresholds below which payments are not made and making payments based on average reductions across days within a month rather than for each day. Settling customer payments at the monthly level – that is, allowing estimated load increases to offset estimated load reductions for all events in the same billing period – greatly improves payment accuracy.
- A more substantive finding is that modifying the program to increase average load impacts per customer has the greatest potential to reduce costs and improve cost-effectiveness. Evidence from various PTR pilots suggests that average load reductions will be higher for opt-in programs than for default enrollment programs. Two scenarios were examined and compared with the current default scenario. The first scenario changes PTR participation from default to opt-in. The second scenario also makes PTR an opt-in program, but assumes that high responders will be targeted and will comprise a large share of participants.
- Figure 1-1 summarizes the findings from this scenario analysis. All three scenarios involve assumptions that produce the same aggregate demand reduction. The left hand axis in the figure shows the amount of total incentive payments that are paid in error. The right hand axis shows the total incentive payments divided by the actual load reduction, expressed in $/kW-yr. The values of $5 million, $1.2 million and $840,000 are the total incentive payments made each year for each scenario. Scenario 1 – default PTR – has costs of $5 million per year, with high error rates and high rebate payments per MW of demand reduction delivered. Scenario 2 – opt-in PTR with no targeting – has incentive costs of $1.2 million per year, with lower error rates and lower costs per MW. Scenario 3 – targeted opt-in PTR – has costs of $840,000 per year, with the lowest errors of all and the lowest cost per kW-yr of benefit. The preferred policy direction is clear: rather than changing baseline methods and payment rules, the best way to improve PTR cost effectiveness is to shift from default to opt-in enrollment and to target customers that deliver large average demand reductions.

Figure 1-1: Scenario Overview



2 Introduction

In 2012, SDG&E implemented one of the two largest default peak time rebate programs in the country (the other being SCE's program). SDG&E's "Reduce Your Use Rewards" program offered approximately 1.2 million residential premises an opportunity to get paid to reduce peak period load based on the following program features:

- The peak period is from 11 AM to 6 PM;
- A Reduce Your Use Day can be called on weekdays or weekends;
- Customers can request to be notified personally the day prior to an event through email or text message.[1] Those who do not enroll in the voluntary notification option can still become aware of events through local media announcements;
- The majority of customers are paid $0.75/kWh of estimated peak period load reduction for each event day and the payment is made as a credit on their bill. Customers with enabling technology that can be controlled by SDG&E, such as customers who are also enrolled in the utility's Summer Saver load control program, were paid $1.25/kWh of reduction.

The amount a customer is paid for each PTR event is based on the difference between metered load and an estimate of what the customer would have used under event conditions if the event had not occurred. This estimated usage is referred to as a baseline. In the Reduce Your Use program, the baseline is estimated for each customer by averaging their electricity usage during the peak period on the highest three out of the prior five, non-event weekdays for an event called on a weekday and the top three out of the five, non-event weekends if a weekend event is called.[2]

PTR was formulated as a no- or low-risk alternative to Critical Peak Pricing (CPP). Both PTR and CPP programs are based on the fact that large scale investment decisions regarding building additional power plants are driven by peak demand during a limited number of critical hours. Reducing demand during these hours reduces supply costs. CPP provides a price discount on most summer days and during non-peak hours on high demand days (event days) in exchange for higher peak period prices on event days. The higher peak period prices provide an incentive to reduce demand. PTR takes the opposite approach. Instead of facing higher prices, customers are offered a rebate if they reduce their demand below a baseline level. They face a similar monetary incentive as in a CPP rate – if they choose to consume, they forgo the rebate – but without the same risk of higher bills if they don't reduce usage on event days.

The "no risk" feature of PTR is one of the primary reasons it is attractive to regulators and other stakeholders. It also allows the program to be implemented on an opt-out basis. Customers that do not reduce their use or receive rebates do not experience substantial changes in their electricity bills (although the cost of payment error, which can be substantial, must be recovered from all ratepayers). In theory, large participation and potentially large aggregate demand reductions could be obtained by placing all customers onto PTR and notifying customers of rebate opportunities either directly or through mass media channels, such as the evening news or community service announcements on the radio.

---

[1] In 2012, roughly 3 to 4% of customers actively enrolled in the voluntary notification program – SDG&E's ex post evaluation of the program showed that measurable savings only came from this group of customers. Customers enrolled in the MyAccount online portal system received PTR e-mail alerts, but they did not actively choose to receive these alerts. MyAccount customers did not show measurable savings.

[2] This is referred to throughout this report as the 3/5 baseline method.

There have been several pilots and implementations of PTR since the concept was first tested by the Anaheim Public Utility District in 2005.[3] These pilots can be split between those that relied on opt-in enrollment in which customers are recruited onto PTR and those that relied on default enrollment. Two key findings from these pilots are:

- In the absence of enabling technology that automates demand response, PTR pilots relying on opt-in enrollment delivered reductions ranging from 7% to 27%;
- Demand reductions for default PTR pilots and full scale implementations are much lower than those for opt-in programs. In studies implemented so far, demand reductions were in the 1 to 3% range.

A PTR program's cost effectiveness and success hinges on the accuracy of the baseline used to settle customer payments and on the magnitude of the demand reductions delivered. Unfortunately, it is very difficult to correctly predict load impacts for individual customers on individual event days using baselines, particularly if customer load impacts are very small, as they usually are for a default program involving the entire population. The large errors produced by an unsuitable baseline can mean that millions of dollars are incorrectly paid to customers who did not actually reduce their load at all. This is compounded by PTR's asymmetric payment structure, which pays customers for estimated load reductions but does not penalize them for estimated load increases.

Many different baseline methodologies have been used by various utilities for a variety of PTR pilots. Little formal work has been published to determine what baseline works best for residential customers enrolled in a PTR program.[4] Most baseline-related studies have been conducted for commercial or industrial demand response programs.[5] However, a baseline that works for customers enrolled in those types of programs may not be suitable for residential PTR customers at SDG&E. The results presented in this report were prepared using a large, randomly selected sample of 2,000 residential customers at SDG&E and are tailored to this program.

The perfect baseline would predict exactly what a customer would have used if they had not participated in a PTR event. This is also known as a reference load or counterfactual. However, no baseline is perfect. Any baseline represents an imperfect estimate of a customer's usage. Nevertheless, some baselines do produce better estimates than others of the counterfactual and do a better job of compensating customers correctly. This report analyzes 21 baseline types that can be used to determine customer payments. Ten different same-day adjustment methods were applied to each baseline, and the analysis was done for the current 11 AM to 6 PM event window, as well as a shorter event window from 1 PM to 6 PM. In total, 420 different baseline/event-window combinations were examined. Some of these baselines conform to SDG&E's baseline implementation requirements, while others do not.[6] Those that conform to SDG&E's implementation requirements can be used by

[3] See Appendix D for a brief summary of these pilots.

[4] Two previous studies conducted on this topic by the FSC Group include:

Oklahoma Corporation Commission Staff Report: Assessment of a Peak Time Rebate Pilot by Oklahoma Gas & Electric Company.

California Public Utilities Commission Application 10-02-028. *Pacific Gas & Electric Company 2010 Rate Design Window Rebuttal Testimony.* Exhibit PG&E-2, Chapter 9. Filed April 3, 2012.

[5] One study completed by the FSC Group was *2011 Statewide Evaluation of California Aggregator Demand Response Programs. Volume II: Baseline Calculations and Accuracy.* http://calmac.org/publications/Aggregator_Statewide_Program_Year_2011_Baseline_Evaluation_Volume_II_.pdf.

[6] These implementation requirements are more fully described in Section 3.

SDG&E with minimal changes to existing operations and IT infrastructure, while non-conforming baselines would be more difficult and costly to implement. Several high-performing baselines are singled out and used to assess the effect of changing payment rules on the cost of payment error. Changes to payment rules are meant to reduce excess payments without changing the baseline methodology.

The remainder of this report is structured as follows. Section 3 describes the study method that was used to evaluate various baseline and incentive payment alternatives. It also includes a detailed description of the baselines studied in this report. Section 4 documents the results of the baseline accuracy analysis while Section 5 examines the effect of changing payment rules on excess payments. Various appendices provide greater detail on selected subjects or additional analysis not discussed in the main body of the report.

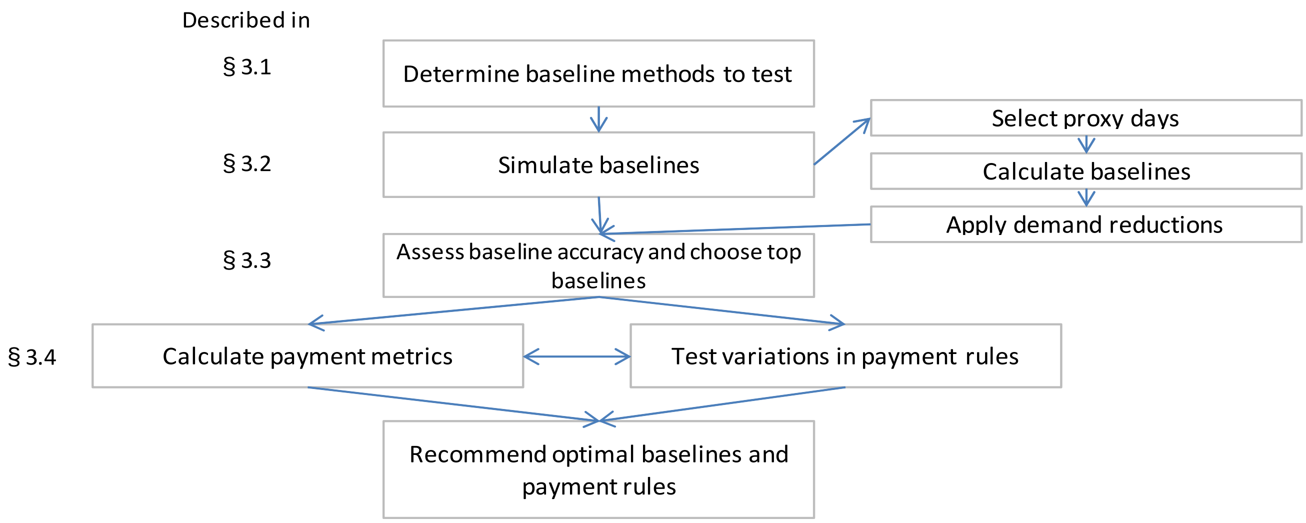
3 Study Methods

As discussed above, this report examines the accuracy of baseline methods for predicting reference loads and load impacts and the accuracy of payments using various baseline methods. While baseline, impact and payment accuracy are related, they are not the same, for reasons explained later in this section.

To assess baseline accuracy, it is necessary to know both the baseline and the true load during the event window. In this study, baseline accuracy is assessed by comparing estimated baselines on days with weather conditions similar to typical PTR event days (referred to as event-like or proxy days) with actual measured load on the same days. A similar process can be used to assess the accuracy of load reductions on the same days by assuming a load reduction on these days and then comparing estimated reductions with the assumed value. This process allows for an examination of baseline and load impact methods under idealized conditions where the correct answer is known.

Figure 3-1 summarizes the process used to complete the analysis objectives. First, a list of baseline methods was developed based on a review of those that have been used elsewhere and logical extensions of those methods. A full list of baselines tested is contained in Section 3.1. Next, for each method, baselines were estimated for a set of proxy days for a large, representative sample of SDG&E residential customers. Proxy days are days with similar weather and system load characteristics to days on which events are typically called. These baselines were compared with the actual load for each baseline method and proxy event day. This process is described in greater detail in Section 3.2. After examining the accuracy and precision of each baseline method, a group of finalists was chosen for further consideration. This method is detailed in Section 3.3; results are summarized in Section 4. This subset of "high performing" baselines was then used to predict load reductions and incentive payments using the current settlement rules as well as alternative payment rules, such as setting threshold amounts below which payments would not be made, to determine if changes in the payment rules would reduce excess payments. This process is described in Section 3.4; results are discussed in Section 5. The final step in the process is to identify those baseline methods and payment rules that have greater impact accuracy and lower payment error compared with the alternatives examined.

Figure 3-1: Evaluation Process



3.1 Baseline Methods Included in Study

The first step in the analysis was to select baseline methods to be assessed. Table 3-1 shows the 21 baseline methods that were evaluated.[7] For each method, 10 in-day adjustments were applied for each of two event windows, the current 11 AM to 6 PM event window and an event window from 1 PM to 6 PM. The latter time period represents the Resource Adequacy window for California. A shorter event window reduces total payments and conforms to Resource Adequacy requirements, but might miss SDG&E's peak hours, which sometimes occur earlier than the peak hours for California's other utilities.

The baseline methods listed in Table 3-1 include ones that meet SDG&E's current implementation criteria and ones that do not. Generally, baseline methods that meet SDG&E's implementation criteria can be adopted with minimal changes to the utility's operations and IT structure. SDG&E's current implementation criteria are:

- Only customer usage data can be used in the baseline calculation (e.g., no weather data);
- The baseline is calculated for each individual event, not averaged across events;
- Interval data is first summed by on-peak and off-peak rate blocks and then the baselines are calculated;
- The baseline should not use kWh data from the day of the event (that is, should not use same-day adjustments);
- Customers should be able to view their bill credits soon after the event (thus precluding using days following an event in the baseline calculation); and
- The baseline should be easy to explain to customers.

[7] The original work scope for this project included a separate assessment of alternative methods for estimating weekend reference loads. However, in light of the findings reported in Section 4 regarding the limited improvement in program performance from modifications to the weekday baseline methods, this additional analysis was unproductive and was dropped.

Non-conforming baseline methods shown in Table 3-1 include weather baselines, which choose days that have similar weather conditions as the event day – and do not consider load – and a regression model that uses both load and weather. Baseline methods shown in Table 3-1 that have been used in other jurisdictions (mainly in conjunction with PTR pilots) identify the relevant jurisdiction in the Source column.

Table 3-1: Baseline Types Included in Evaluation

Baseline Category	Baseline	Description	Source
Meet SDG&E Implementation Criteria	3/3	Average 3 of last 3 eligible[i] days	Wisconsin Energy
	3/3, weighted	Use 3 of last 3 eligible days; weight days 50%, 30% and 20% respectively; more recent days receive higher weight	-
	3/5	Average the top 3 of the last 5 eligible days	SDG&E
	3/5, weighted	Use top 3 of the last 5 eligible days; weight days 50%, 30% and 20% respectively; more recent days receive higher weight	-
	3/5, +5%	Average 3 of last 5 eligible days and adjust upward by 5% for all customers	Ontario
	4/5	Average top 4 of the last 5 eligible days	-
	5/5	Average top 5 of the last 5 eligible days	-
	3/10	Average top 3 of the last 10 eligible days	PJM
	5/10	Average top 5 of the last 10 eligible days	New York ISO
	10/10	Average 10 of the last 10 eligible days	-
	3/20	Average top 3 of the last 20 eligible days	PowerCents DC
	5/20	Average top 5 of the last 20 eligible days	-
	10/20	Average top 10 of the last 20 eligible days	-
Do Not Meet SDG&E Implementation Criteria	Weather-matching - 3 days	Average 3 days with similar weather during the last three months	-
	Weather-matching - 4 days	Average 4 days with similar weather during the last three months	-
	Weather-matching - 5 days	Average 5 days with similar weather during the last three months	-
	3/14 with THI	Average top 3 of last 14 eligible days (including weekends); discard days that don't have similar weather based on temperature-humidity index (THI)	BG&E
	Regression	Regress energy use during event window on weather and day of week[ii]	-
	Maximum Temp. Weather Baseline	Assign days with high temperatures exceeding 80°F to 1 of 3 bins based on maximum temperature; baseline equals the average peak-period load on non-event days that have maximum temperature values for the relevant bin corresponding to the event day	-
	CDD Weather Baseline	Assign days with high temperatures exceeding 80°F to 1 of 3 bins based on CDD[iii] for the day; baseline equals the average peak-period load on non-event days that have CDD values for the relevant bin corresponding to the event day	-

Baseline Category	Baseline	Description	Source
	Sum CDH Weather Baseline	Assign days with high temperatures exceeding 80°F to 1 of 3 bins based on the total CDH[iii] for the day; baseline equals the average peak-period load on non-event days that have CDH values for the relevant bin corresponding to the event day	-

[i] An eligible day is defined as a day preceding the event that is a non-event, non-holiday weekday, except for 3/14 with THI, where an eligible day is defined as a preceding non-event, non-holiday day (*including* weekends).

[ii] The regression can be expressed as follows:

$$eventkwh_{i,d,s} = \alpha + \beta_1 * sumCDH_{i,d,s} + \beta_2 * morningCDH_{id,s} + \beta_3 * DOW_d + \beta_4 * Month_d + \beta_5 * Year_d + \epsilon$$

where α is a constant, $\beta_1 - \beta_5$ are coefficients, ϵ equals the errors, sumCDH equals the total CDH during the event, morningCDH equals the CDH in the morning of the event, DOW is a fixed effect for the day of week, month is a fixed effect for the month of the year, and year is a fixed effect for the year. The regression is estimated for each customer (i) and event period (s) using non-event data; then, the coefficients are applied to each event day (d).

[iii] CDH and CDD are cooling degree hour and cooling degree day, respectively. A cooling degree hour is defined as the maximum of the temperature minus a base value or 0. For example, if the temperature in a given hour is 90 degrees, the cooling degree hour value for that hour is 90-70=20. If the temperature is 60 degrees, the cooling degree hour value for that hour is max(60-70, 0), or 0. A cooling degree day applies the same procedure to the average temperature during the entire day.

3.2 Simulation Approach

In order to assess baseline accuracy, both the estimated baseline and the true load (what actually happened) must be known. The simulation approach ensures that we have both elements needed to make a comparison by calculating baselines for event-like days (days that could have been events, but were not) and then comparing those baselines to the actual load. Using this method, the true impact and the impact estimated by the baseline is known, allowing us to determine how accurately each baseline estimates load impacts.

To implement the simulation, we:

1. Select multiple proxy event days: 15 proxy event days were selected that match, as closely as possible, actual event conditions observed when SDG&E's demand response and dynamic pricing programs are activated. To be eligible for events, two conditions must be met. First, the day-ahead system load must be projected to exceed 3,600 MW. Second, the forecasted peak temperature at Miramar must exceed 85°F. The current criteria do not automatically trigger events, and as a consequence, PTR is not dispatched on many eligible days. The 15 proxy days occured across 3 years (2010 through 2012), yielding an average of 5 event days per year. The dates, system load, and weather conditions for the 15 proxy event days are listed in Appendix A.
2. Calculate the baselines and load estimaed impacts using each of the baseline methods listed in Table 3-1: Baselines were calculated for each event day and for a randomly selected sample of 2,000 SDG&E residential customers.
3. *Apply demand reductions to the unperturbed load:* For each event day, assumed demand reductions were subtracted from the unperturbed loads. These demand reductions equal 0%, 2%, 10% and 20% of the reference load to simulate a range of average customer responses that might occur for different program implementation scenarios or customer segments. The higher assumed reductions would be more likely to occur from an implementation in which a large percent of participants had PCTs and/or signed up for personal notification while the lower average reductions are more consistent with SDG&E's current default implementation.

All three steps outlined above were implemented. However, the selection of "high performing" baseline methods to be used to assess payment accuracy was based on a comparison of estimated baselines with actual load, rather than on the basis of estimated load reductions with assumed load reductions. The estimated load reductions were used in the assessment of payment accuracy.

3.3 Measuring Baseline and Impact Accuracy

Once baselines are calculated and demand reductions have been applied, it is possible to compare how accurate each baseline is compared to the actual load without PTR, and how accurate impacts estimated by the baseline are compared to true impacts (which are introduced artificially and therefore known). This section presents background material on FSC's approach toward measuring baseline accuracy and introduces the baseline accuracy metrics used throughout this report.

3.3.1 Baseline Error and Impact Error

Before discussing the various accuracy metrics used in this study, it is important to distinguish between the accuracy of the baseline (compared to the actual load) and the accuracy of impacts calculated by the baseline (compared to impacts calculated using actual load). These are two related but distinct types of accuracy. For example, assume that in a simulation exercise similar to the one described in Section 3.2 a customer consumes 2.0 kWh and the baseline estimates that they consume 2.2 kWh. The percent in the baseline can be calculated as follows:

$$\frac{Baseline}{Actual\ Load} - 1 = \frac{2.2\ kWh}{2.0\ kWh} - 1 = +10\%$$

The baseline overestimates true usage by 10%. Suppose now that the customer reduces their load by 10%, or 0.2 kWh. Under this assumption, the metered load would equal 1.8 kWh (2.0 – 0.2) and the estimated load reduction would equal 0.4 kWh (2.2 – 1.8). Using actual load, the impact is correct, and is 0.2 kWh. However, using the baseline, the impact is 0.4 kWh (2.2 kWh – 1.8 kWh). Thus, the impact error equals:

$$\frac{Estimated\ Impact}{True\ Impact} - 1 = \frac{0.4kWh}{0.2kWh} - 1 = +100\%$$

The impact error in percentage terms is ten times as large as the baseline error. In fact, the smaller the true load reduction as a fraction of the customer's load, the more baseline error is magnified in the estimated demand reductions. While a relatively small baseline error, such as 2 or 3%, might be perfectly acceptable, it is important to realize that this error will balloon when expressed in terms of the impact, particularly when that impact is small (as it usually is when an entire population participates by default). The relationship between impact error and baseline error is expressed as follows:

$$\frac{Baseline\ Error\ (\%)}{Impact\ (\%\ of\ true\ load)} = Impact\ Error\ (\%)$$

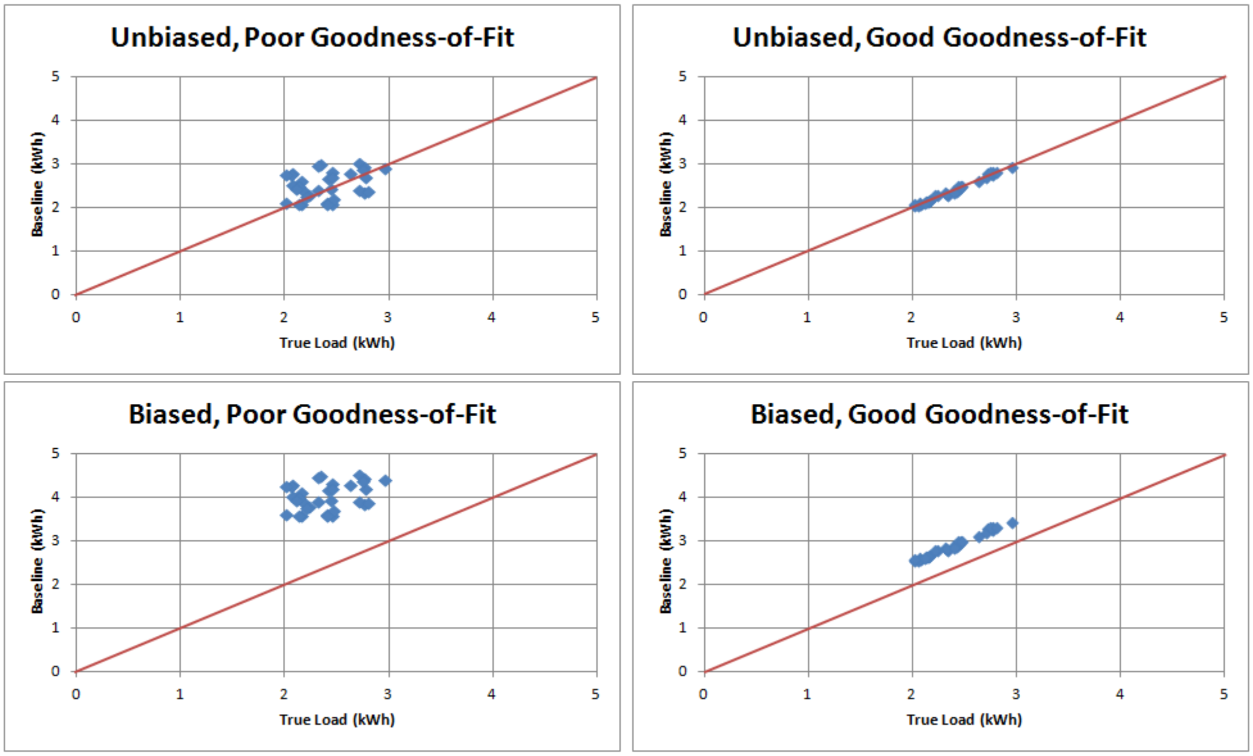
For the sake of simplicity, and to avoid the need to present results for many different impact sizes, the discussion of high performing baselines in Section 4 focuses on baseline error rather than impact error, even though the sole purpose of baselines is to estimate demand reductions. Because baseline error and impact error are closely related, and because baseline error does not vary with impact size, either variable could have been used to screen baseline methods, but the baseline comparison was simpler to present. However, when assessing payment accuracy, impact error is the driving force, and both impact error and payment error vary with impact size. The relationship between payment error and impact size is complicated due to the asymmetric nature of PTR incentives. For small average impacts, payment error is a greater percent of total program incentive payments than when average impacts are larger.

3.3.2 Bias and Goodness-of-fit

A well-performing baseline should be unbiased and precise (have high goodness-of-fit). As shown in Figure 3-2, a biased baseline, on average, overestimates or underestimates the true load. In the "biased" panes of the graph, all dots are consistently above the red line that reflects an accurate baseline.[8] An unbiased baseline does not overestimate or underestimate the true load on average. The two baseline methods depicted at the top of Figure 3-2 are unbiased. Bias is distinct from precision, or goodness-of-fit, which refers to the amount of spread or variance around the average estimate. The two baseline methods shown on the right hand side of Figure 3-2 have better precision than the two on the left. Thus, as shown in the figure, a baseline can be unbiased and have poor goodness-of-fit, meaning that, on average, the baseline is correct, but that it is not necessarily correct for individual customers on individual days. Such a baseline has a wide spread of errors, but the errors cancel each other out. Obviously, a good baseline method would be both unbiased and highly precise.

[8] Bias can also be negative, in which case the dots would be below the line in the figure. Also, not all dots need to be above or below the line to create an upward or downward bias for the program overall. If, for example, a baseline was upwardly biased for large users and downwardly biased for small users, the overall bias could go either way depending on the usage weighed average of the biases.

Figure 3-2: Bias and Goodness-of-fit



3.3.3 Baseline Accuracy Metrics

As noted above, the best baselines are both unbiased and highly precise. Ideally, a baseline would perform well in aggregate at the program level as well as for individual customers. This distinction is important because some customers use much more electricity than others. When program level accuracy is considered, larger customers are weighted more heavily than smaller customers since they account for a larger share of the load, baseline error and PTR payments.

Table 3-2 shows the metrics used to evaluate the baselines shown in Table 3-1. These metrics test whether a baseline minimizes bias and maximizes goodness-of-fit for individual customers. The analysis of payment rules was only conducted for baselines that performed relatively well on the metrics shown in Table 3-2.

Table 3-2: Baseline Accuracy Statistics

Statistic Type	Statistic Level	Statistic	Formula	Description	Typical Values
Bias	Individual Customer	Average Percent Error (Distribution)	$\frac{\frac{1}{n}\sum \hat{y}_{i,t}}{\frac{1}{n}\sum y_{i,t}} - 1$	Calculated for each individual customer. We show the distribution of bias across individual customers.	Expressed in percentage terms. Can be positive or negative. The closer to zero, the better.
	Program	Average Percent Error	$\frac{\sum \hat{y}_{i,t}}{\sum y_{i,t}} - 1$	Sums up baseline and actual values for individual customers and event days for the entire program; calculates error statistic from these values.	Expressed in percentage terms. Can be positive or negative. The closer to zero, the better.
Goodness-of-fit	Program	Absolute Sum of Errors	$\sum_{i=1}^{n} \lvert \hat{y}_{i,t} - y_{i,t} \rvert$	Sums up absolute errors for individual customers and event days.	Expressed in kWh terms. Can only be positive. The smaller the number, the better. Equivalent to summing up the distances between the dots and the red line in Fig. 3-2.

The statistics in this table use the following nomenclature:

- y - actual kWh
- $\hat{y}$ - kWh estimated by the baseline
- $_{i,t}$ – individual customers and events
- n - total number of event days

3.4 Payment Rules Analysis

After a number of baselines were selected that perform relatively well on tests of bias and goodness-of-fit, payment accuracy was assessed. This section describes the tested payment rules and discusses the metrics used to determine payment accuracy. Before discussing those metrics, the next section introduces a key concept needed to understand the asymmetry of PTR payments.

3.4.1 Payment Asymmetry and Impact Size

As discussed previously, PTR payments are asymmetrical, meaning that customers are paid when the baseline erroneously estimates a decrease in load, but they are not asked to pay the PTR incentive amount per kWh when the baseline erroneously measures an increase in load. This asymmetry is illustrated in Figure 3-3, which depicts a scenario in which it is assumed that customers do not respond to an event. We also assume a baseline method that is completely unbiased for all customers for the average event but has small errors for individual event days.[9] The graph shows estimated impacts for individual customers on one event day (not for the average day). The true impact (0 kWh) is depicted by the green line. The blue circles and red Xs show the estimated impacts for individual customers. Customers shown in red do not get paid because the baseline erroneously

[9] For individual event days the baseline error is less than ±10% for 80% of customers. As shown later, this baseline is more than 5 times more precise than any of the baselines tested.

indicates an increase in demand but customers in blue do get paid because the baseline erroneously indicates a decrease in their load. In reality, no one is supposed to get paid and all payments made to customers are due to measurement error. As seen, despite assuming a highly accurate baseline at the individual customer level (which is not achievable by any baseline that was tested), 50% of customers receive a payment even though no load reductions were provided. Despite using a highly accurate baseline that is not even achievable in reality, the asymmetry in payment rules leads to a large overpayment.

Figure 3-3: Payments to Individual Customers Assuming No Impacts and an Unbiased Baseline Method at the Individual Customer Level

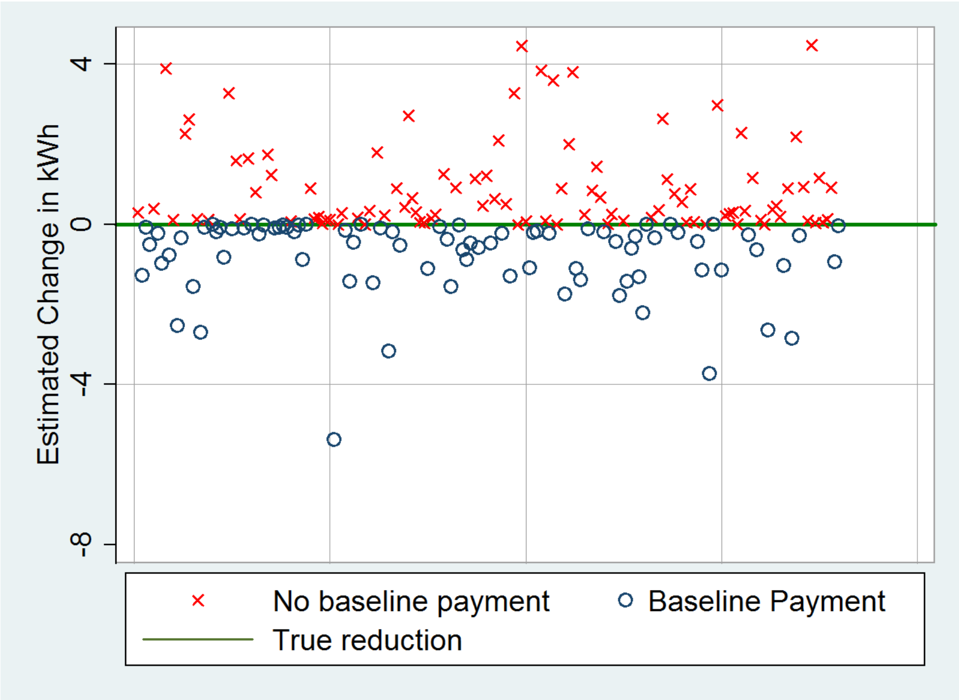


Figure 3-4 shows an alternate scenario using the same day, same set of customers and same ideal baseline. The only difference is that under this scenario, we assume that each customer reduces demand by an average of 10%. The green line shows the average customer's demand reduction. Again, blue circles represent customers on an individual day that get paid, while red Xs represent customers that do not get paid. In this scenario, all customers are supposed to get paid. In reality, some customers get paid while others do not. Some customers are underpaid while others are overpaid but in this case, overestimates are mostly canceled out by underestimates and excess payments would be much smaller than in the first scenario. In practice, payment error will be larger because baselines are far less accurate than the baseline simulated here and it is implausible that every single customer will reduce demand by 10%. However, the example illustrates the substantial role that payment asymmetry plays in payment error when demand reductions are relatively small.

Figure 3-4: Payments to Individual Customers Assuming 10% Impacts and an Unbiased Baseline Method at the Individual Customer Level

3.4.2 Alternative Payment Rules

The current rule used by SDG&E to settle customer payments takes the load reduction estimated using the 3/5 baseline, rounds it to the nearest 1 kWh, and pays $0.75/kWh of estimated load reduction. The rules do not charge customers for load increases estimated by the baseline. As discussed above, this last point is an inherent feature of PTR programs and complicates the search for a baseline and set of payment rules that, ideally, pay people accurately but more realistically at least reduce the amount of payment made for non-existent load reductions. Changes in payment rules, such as setting minimum thresholds below which payments are not made, have the potential to reduce payments made in error for most baselines.

An assessment of payment error was done for the following payment rules, in addition to the current rule:

- **Threshold – relative.** This scheme assumes a payment of $0.75/kWh if the estimated demand reduction is greater than 10% of the baseline. For example, if the estimated reduction is 1 kWh but that 1 kWh equaled only 5% of the baseline load, the payment would equal $0. On the other hand, if that 1 kWh equaled 20% of the baseline load, the payment would be $0.75/kWh.
- **Threshold – absolute.** This scheme is very similar to the first rule, but the threshold would be set in absolute terms rather than in relative terms. For example, if the threshold equaled 2 kWh and the estimated demand reduction equaled 1 kWh, the payment would be $0. On the other hand, if the estimated reduction equaled 3 kWh, the payment would equal $2.25 (3 kWh * $0.75/kWh).
- **Graduated – relative.** This scheme assumes a payment of $0.50/kWh for estimated demand reductions between 0 and 20% of the baseline, and $1.00/kWh for estimated demand reductions greater than 20% of the baseline. Thus, if the estimated reduction equaled 3 kWh and this amounts to 30% of the baseline load, the first 2 kWh would be paid at $0.50/kWh

while the last kWh would be paid at $1.00/kWh. The total payment in this example would equal $2.00.

- **Graduated – absolute.** In this scheme, incentives would be paid at a rate of $0.50/kWh for estimated demand reductions less than 2 kWh and at a rate of $1.00/kWh for estimated demand reductions greater than 2 kWh. In this case, if the estimated demand reduction equaled 3 kWh, the first 2 kWh would be paid at $0.50/kWh while the final kWh would be paid at $1.00/kWh, for a total payment of $2.00.
- **Monthly.** This payment scheme adds up all estimated peak-period load reductions, including load increases as well as decreases, for an entire calendar month before calculating payments, and then pays $0.75/kWh for the net estimated demand reduction across all events in the month. For example, if there were two event days in one month, and the baseline led to an estimated load reduction of 3 kWh on one day and increase of 2 kWh on the other day, the incentive would pay the customer for the average reduction of 1 kWh (3 kWh – 2 kWh = 1 kWh). If, on the other hand, the baseline found an increase of 3 kWh on one day and a decrease of 2 kWh on the other, the customer would not have to pay for the increase, he or she would not receive any payment in that month (2 kWh – 3 kWh = -1 kWh, and customers are not paid for load increases).

3.4.3 Payment Accuracy Metrics

As with the baseline analysis, we assess payment accuracy based on a variety of metrics. As before, some statistics apply to the program level and weight larger customers more heavily, whereas others are at the individual customer level and treat all customers equally. Table 3-3 shows the metrics used to evaluate payment accuracy for the handful of well-performing baselines that were selected for additional study.

Table 3-3: Payment Accuracy Statistics

Statistic Level	Statistic	Formula	Description	Typical Values
Program	Total Error ($)	$\sum(\widehat{\$}_{i,t} - \$_{i,t})$	Sums up total payments made in error for each individual customer and event day.	Expressed in dollars. The greater the value, the more the payment error.
	% Excess Payment	$\frac{\sum(\widehat{\$}_{i,t} - \$_{i,t})}{\sum \widehat{\$}_{i,t}}$	Divides total payment error by the amount that was paid out to all customers.	Expressed in percentage terms. Ranges from 0 to 100%, with higher numbers being worse.
	$/kW-year	$\frac{\sum \widehat{\$}_{i,t}}{\frac{1}{n}\sum kW_{i,t}}$	Represents a capacity value for PTR reductions. Takes total estimated payment and divides by the total kW actually delivered by the program on the average event day. It does not factor in cost-effectiveness adjustments (e.g., maximum event duration) or program administration costs.	Expressed in dollar terms, with higher numbers being worse.
Individual Customer	% under/ over/ correctly paid	$\frac{\sum(\widehat{\$}_i < \$_i)}{n}; \frac{\sum(\widehat{\$}_i > \$_i)}{n}; \frac{\sum(\widehat{\$}_i = \$_i)}{n}$	Counts the number of customers who were over/under/correctly paid and divides this by the total number of customers.	Expressed in percentage terms. Ranges from 0 to 100%. For over/under payments, values close to 0 are better. For "correct payments" values closer to 100% are better.
	% Error (Distribution)	$\frac{\frac{1}{n}\sum \widehat{\$}_{i,t}}{\frac{1}{n}\sum \$_{i,t}} - 1$	Shows a distribution of payment errors for individual customers on the average event day.	Expressed in percentage terms, with values closer to 0 being better.

4 Baseline Accuracy Results

This section presents an assessment of accuracy for the baselines listed in Table 3-1 using the metrics described in Section 3. Results are presented at the program level and at the individual customer level. Program level statistics are calculated after adding up the baselines and actual load for all customers in the sample, while individual level results are calculated individually for each customer. The program level approach tends to weight customers who use more electricity more heavily because all loads are added together before calculating any statistics.

Because the analysis was done for a large number of baselines, results are shown in graphical format. Appendix B includes tables with source values for all graphs shown in this section of the report. In addition, detailed interactive tables were provided to SDG&E with this report. This section also includes selected results for the analysis of same-day adjustments and concludes with a recommendation for the five highest performing baseline methods that are used as the basis for the assessment of payment accuracy under alternate payment rules in Section 5.

4.1 Program Level Results

This section presents program level baseline accuracy results. Section 4.2 presents results at the individual customer level. As discussed previously, it is much easier to find a baseline method that works well on average across all customers and all event days than it is to find one that works well at the individual customer level for each event day, although some do perform marginally better than others.

4.1.1 Bias

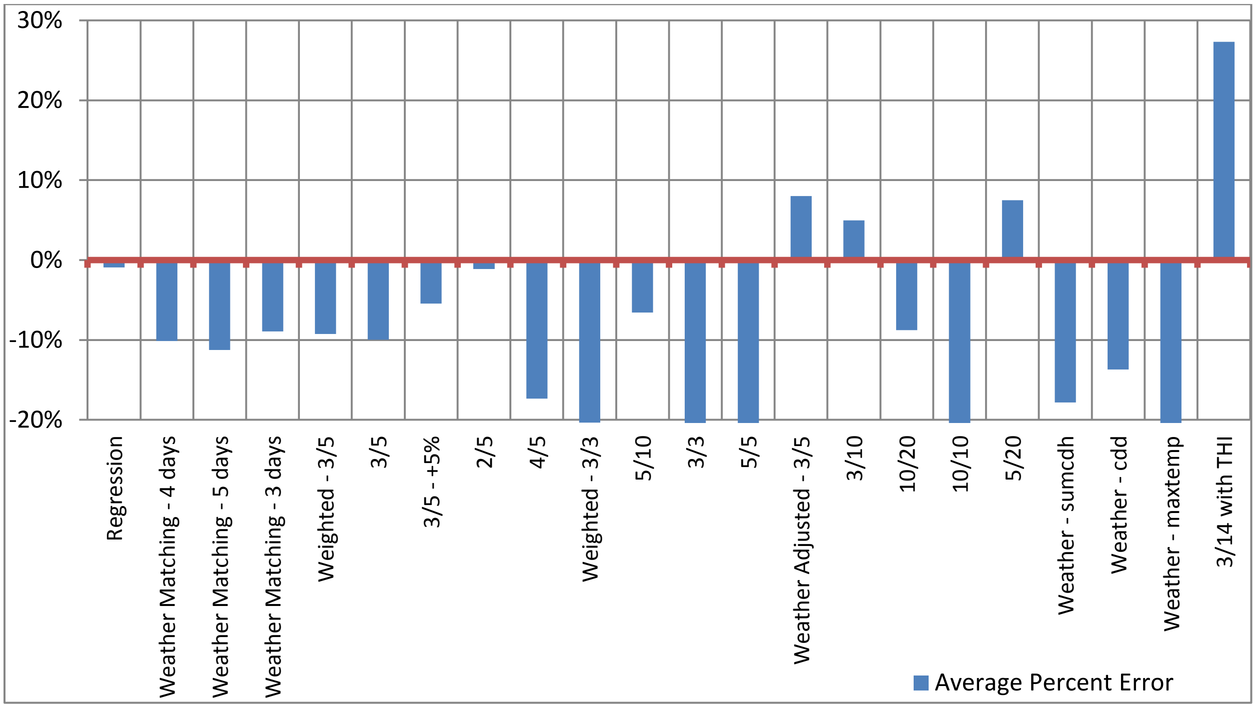
Figure 4-1 presents the average percent error at the program level for the 21 weekday baseline methods listed in Table 3-1.[10] For clarity, the red line marks an average percent error of zero. Baselines closer to the red line exhibit less program-level bias. Recall that the ideal baseline should exhibit zero bias because baseline bias increases dramatically when translated into impact bias, particularly when impacts are relatively small (see section 3.3.1).

Except for the regression and the 2/5 baseline, all baselines exhibit a significant amount of baseline bias. Most other baselines, including the current 3/5 baseline, are biased downward, indicating that they tend to underestimate the actual load and thus underestimate the average customer's load reduction. By extension, if a customer actually reduces their load during an event, a downwardly biased baseline underpays that customer. The only baseline methods that overestimate load reductions for the average customer are the 3/10, 5/20, and 3/14 with THI baselines.

The logic for why almost all baselines shown in Figure 4-2 are downwardly biased is intuitive. PTR events are called on days when SDG&E anticipates a shortage in capacity. These days tend to be much hotter than the average day. Because of increases in the use of air conditioning, most customers use much more electricity on these event days than they do ordinarily (hence the shortage in capacity). However, the days used to calculate the baseline cannot include those event days. Customers tend to use less electricity on non-event days. When comparing a baseline calculated using relatively cool, low-load non-event days to the actual load on a hot, high-load event day, it is logical that the baseline would be lower than the event day, and that it would be biased downward.

[10] Baselines are ordered by increasing sum of absolute errors, which is shown Figure 4-2.

Figure 4-1: Average Percent Baseline Error for Each Baseline Method



4.1.2 Goodness-of-fit

Figure 4-2 presents results for program-level goodness-of-fit using the sum of absolute errors.[11] The sum of absolute errors simply adds up the absolute value of errors for individual customers and individual event days. In this figure, lower bars indicate that a given baseline exhibits better program-level goodness-of-fit (or precision). The baselines performing best using this metric are the regression,[12] the three weather-matching baselines and several variations of the 3/5 day-matching baseline. The un-weighted, unadjusted 3/5 day-matching baseline is the one currently used by SDG&E to settle customer payments. Ranking near the bottom are weather-bin baseline methods and ranking at the very bottom is the 3/14 with THI baseline.

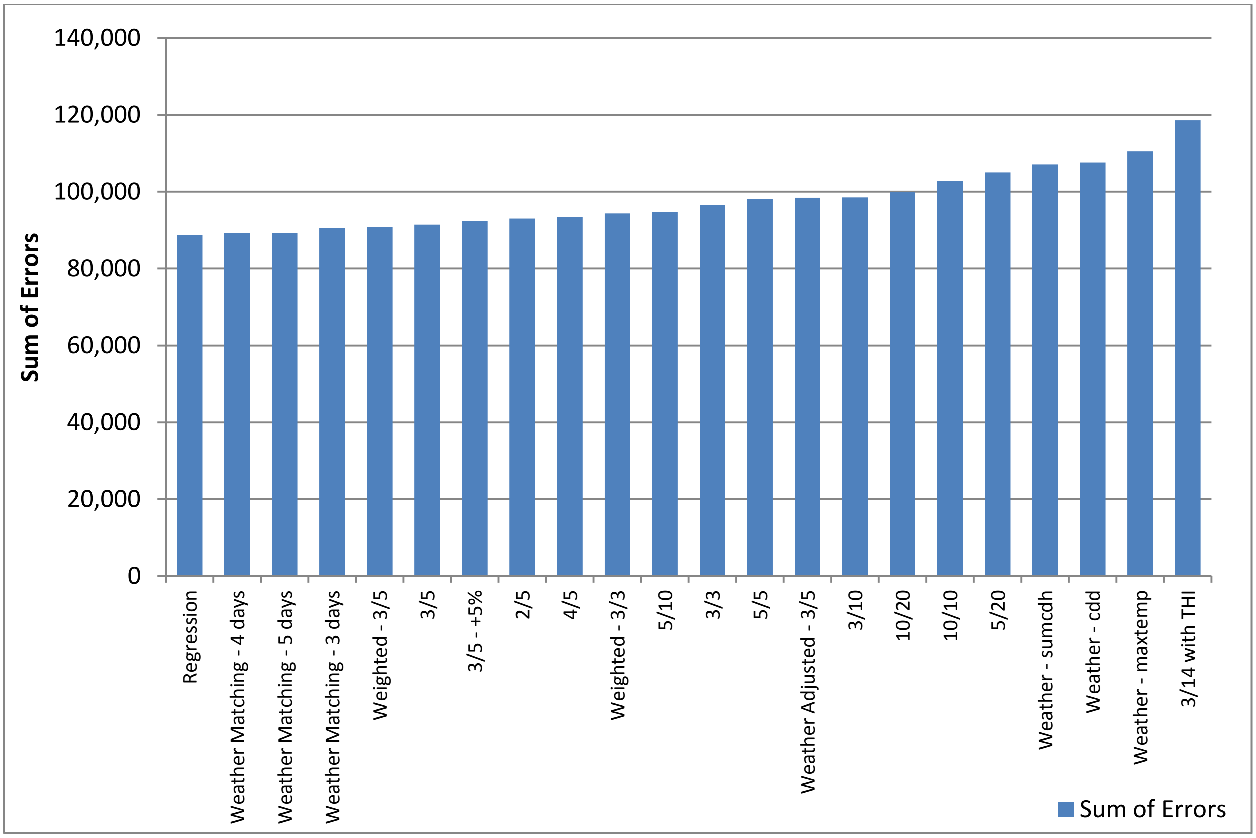
It should be noted that the poor performance by the weather-bin methods may largely be due to the particular binning method used and the time-period over which the binning occurred. These particular baselines included data over three years and segmented all eligible days into three bins for days with maximum temperatures above 80°F. Creating bins based on days that did not go back as far as three years, increasing the number of bins, setting the temperature threshold higher, or other changes could change the performance of a baseline that uses a weather data binning approach.

---

11 The sum of errors squared is also shown in Appendix B.

12 The regression also far outperforms other baselines on the sum of errors squared test. This is because, by design, a regression attempts to minimize the sum of errors squared. The sum of errors squared test is very similar to the sum of absolute errors test; instead of summing the absolute value of errors, it sums the square of errors.

Figure 4-2: Sum of Absolute Errors by Baseline Method



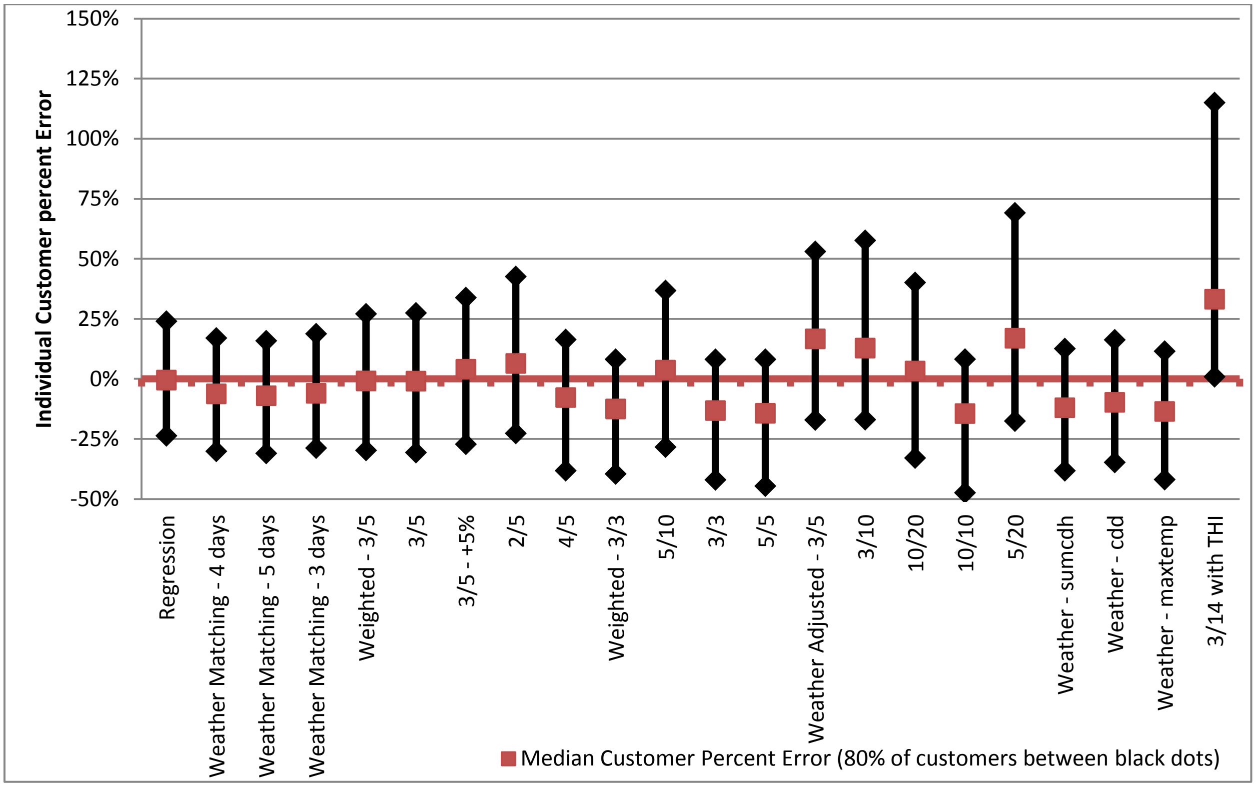
4.2 Individual Customer Results

Section 4.1 examined bias and precision at the program level. This section focuses on these metrics for individual customers. In general, there is considerably more error for individual customers across multiple event days than for the average customer.

Figure 4-3 shows the distribution of bias across customers for each baseline method. The average percent error was calculated for each customer across 15 proxy event days over three years. In the figure, the red dot represents the median value for the distribution of individual bias estimates and the end-points of the black lines represent the 10th and 90th percentiles for this distribution. Thus, 80% of all customers have average percent errors that are between the black dots, and 20% of customers have bias metrics that lie outside those boundaries. For example, for the regression baseline method, 80% of customers have average biases that are within ±25% of the true reference load, and 20% of customers have biases that are more the 25% higher or lower than the true reference load. If the red dot is on 0, it means that the baseline for the median customer has no bias across the 15 proxy days.[13]

[13] A baseline method with a median individual customer bias of 0% does not guarantee that the method is unbiased at the program level. This would only be true if the distribution of individual biases was perfectly symmetrical and if bias was completely random across customers with different usage levels. Even if the distribution of individual customer biases were symmetrical, the program level bias would not be 0% if, for example, larger customers were more likely to have upwardly biased baseline values.

Figure 4-3: Distribution of Individual Customer-level Bias

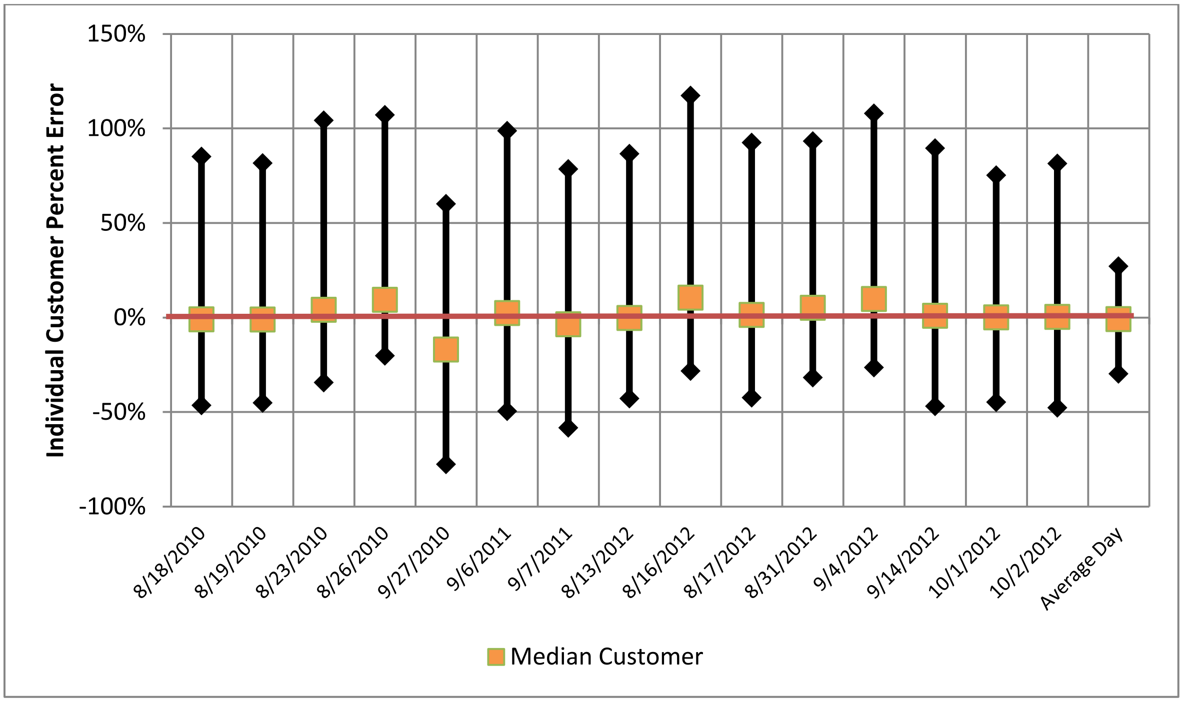


As seen in Figure 4-3, every baseline method except one has a median error that is within ±25% of 0. The notable exception is the 3/14 with THI baseline method, which has a very skewed distribution; its median value is roughly 35%, and 10% of customers have baseline error greater than 125% of the true reference load. The regression and 3/5 baselines do very well, as do the weather-matching baselines. Several baselines show little bias for the median customer, but all of them produce estimates that are more than 25% off for at least 20% of customers. Many are biased to a much greater extent for a larger percent of customers.

It is important to keep in mind that these distributions are based on 15 proxy days over three years. If these statistics were calculated using fewer event days, or shown for individual years, the errors would be even greater than the relatively poor performance depicted in Figure 4-3. Moreover, for an individual event day, customer level error distributions will be much wider, as shown in Figure 4-4.

Figure 4-4 shows the distribution of customer-level errors by event day for the 3/5 baseline method. The average event day is shown at the far right of the figure. This value is the same as that for the 3/5 baseline in Figure 4-3. Figure 4-4 shows that the error on individual days is much greater than the error on the average event day. While the median customer error is fairly constant and centered near 0% on most days, the error range that includes 80% of customers (black bars) is very broad, equaling or exceeding +100% on five days and -50% on three days. Hotter days tend to coincide with underestimates of the true load, while cooler days tend to produce overestimates of the true load.

Figure 4-4: Distribution of Baseline Errors across Customers for Individual Event Days for the 3/5 Baseline Method



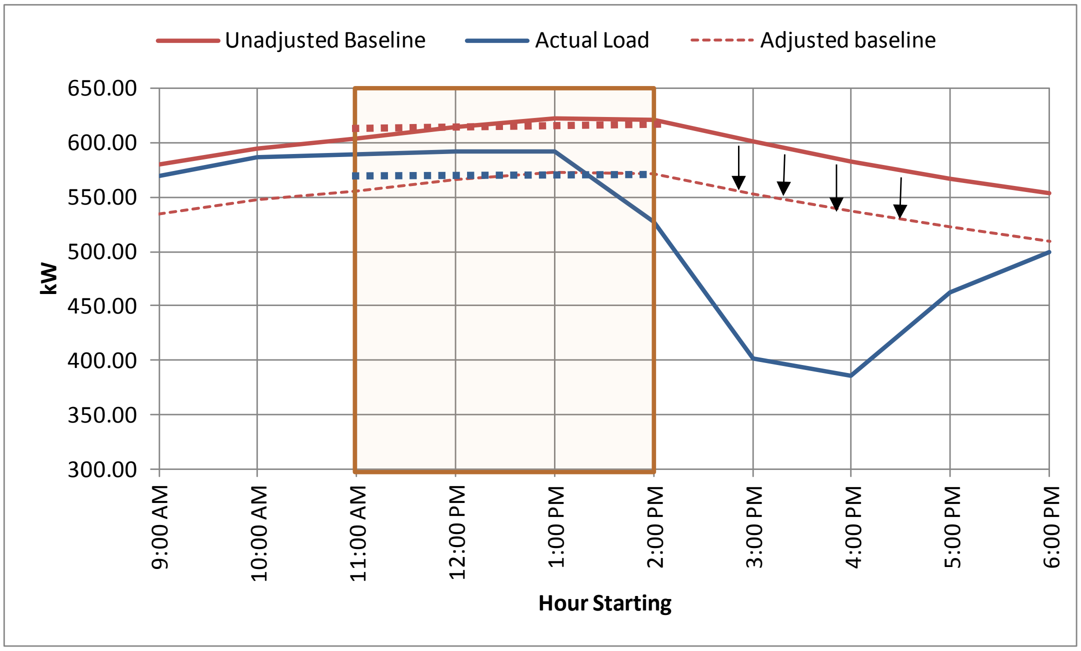
4.3 Same-day Adjustments

One possibility for improving baseline accuracy is to adjust initial baselines by using load on the event day just before the event window. The logic of this is that load just before the start of the event is a good predictor of load during the event, and taking it into account will improve the accuracy of reference loads compared to just using loads from prior days. This process is known as *same-day adjustment*.

Same-day adjustment modifies the initial reference load estimate by comparing load in the hours leading up to the event window on the event day with load during the same hours on the prior days that were used to develop the preliminary reference load estimate. For example, if the pre-event baseline is above the actual load before the event, the initial baseline value is adjusted downward. Alternatively, if the load during pre-event hours is less than the actual load, the initial estimate is adjusted upward. It is important to note that same-day adjustment procedures implicitly assume that differences between the baseline and actual loads during hours leading up to an event are due to predictive error and not due to customer behavior, such as pre-cooling a house or implementing demand reductions early by adjusting thermostats in the morning prior to leaving for work. If these logical behaviors occur, same-day adjustments could increase baseline error, perhaps dramatically.

Figure 4-5 illustrates the baseline adjustment process. In the example, the event starts at 3 PM. The first three of the four hours leading up to the event, from 11 AM to 2 PM, are used to calculate the adjustment. The blue line represents the actual load for the day. The red line reflects the calculated baseline prior to the application of same-day adjustments. In this example, in the hours leading up to the event, the unadjusted baseline is higher than the actual load. The baseline adjustment process assumes this difference is due to error. To correct for this difference, the baseline is calibrated downward by roughly 8%, as reflected by the red dotted line.

Figure 4-5: Example of Baseline Same-day Adjustment Process



$$Adjustment = \frac{Avg.kW\ during\ adjusment\ period}{Avg.unadjusted\ baseline\ over\ adjustment\ period} = \frac{571}{619} = 92.2\%$$

Same-day adjustments are often capped because doing so can reduce the high baseline error and variance resulting from unlimited adjustments. Adjustments are also capped because unlimited adjustments can introduce the potential for manipulation of pre-event loads to bias baselines. The concern is that participants might game the system by increasing their electricity use during the adjustment period, leading to baselines that are too high and, thus, overestimate actual demand reductions.[14] Capping the magnitude of the adjustment limits the potential for this kind of abuse.

The same-day adjustments assessed here use a two-hour window, from 9 to 11 AM, as the adjustment period. A total of 10 same-day adjustment methods were tested, including unadjusted baselines, uncapped baselines, upward/downward caps of 1.1x through 1.4x, and upward only adjustments capped at 1.1x through 1.4x. These adjustment caps are expressed as ratios. A cap of +/-1.4x means that the adjustment ratio cannot be larger than 1.4 or smaller than 1/1.4 (approximately 0.71), whereas a cap of +1.4x means that the adjustment ratio cannot be larger than 1.4 or smaller than 1.

Figure 4-6 shows the effect of applying same-day adjustments to four relatively well-performing baselines using the sum of the absolute errors as the performance metric.[15] This is the same

[14] This analysis does not attempt to determine whether manipulation of baselines has taken place or will occur in the future and takes no position on this issue. Moreover, it is impossible to distinguish between gaming to manipulate the baseline and pre-cooling before an event, which is perfectly legitimate behavior.

[15] The regression, which performed best on nearly all baseline accuracy metrics, was not included in this list because it already uses some pre-event variables and the added benefits of adjusting it were perceived to be small. In addition, to calculate an adjustment, a baseline needs to make a prediction for pre-event hours, which this regression was not designed to do.

precision metric discussed in Section 4.1.2. The unadjusted baseline is shown in red. All four of these baselines benefit from some minor adjustments but these benefits recede as the adjustment cap is loosened. Removing the adjustment cap altogether and allowing adjustments to be unlimited increases the absolute error dramatically for all baselines. Another interesting pattern is that upward-only adjustments (e.g., +1.1x) generally do not perform as well as up-and-down adjustments (e.g. +/-1.1x).

Figure 4-6: Sum of Absolute Error by Baseline Type and Adjustment Cap

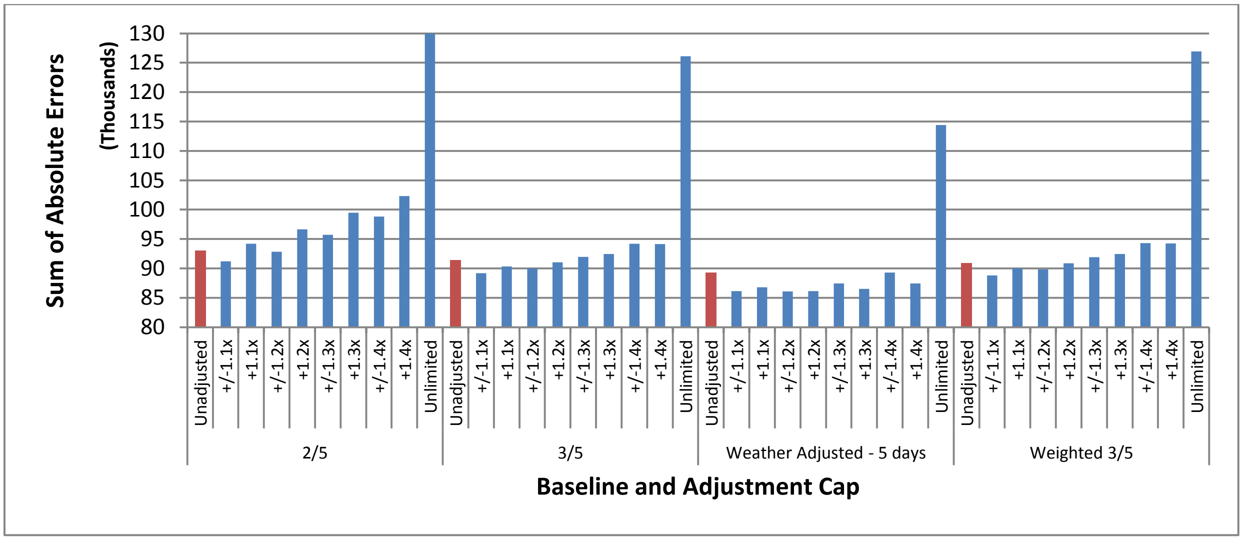
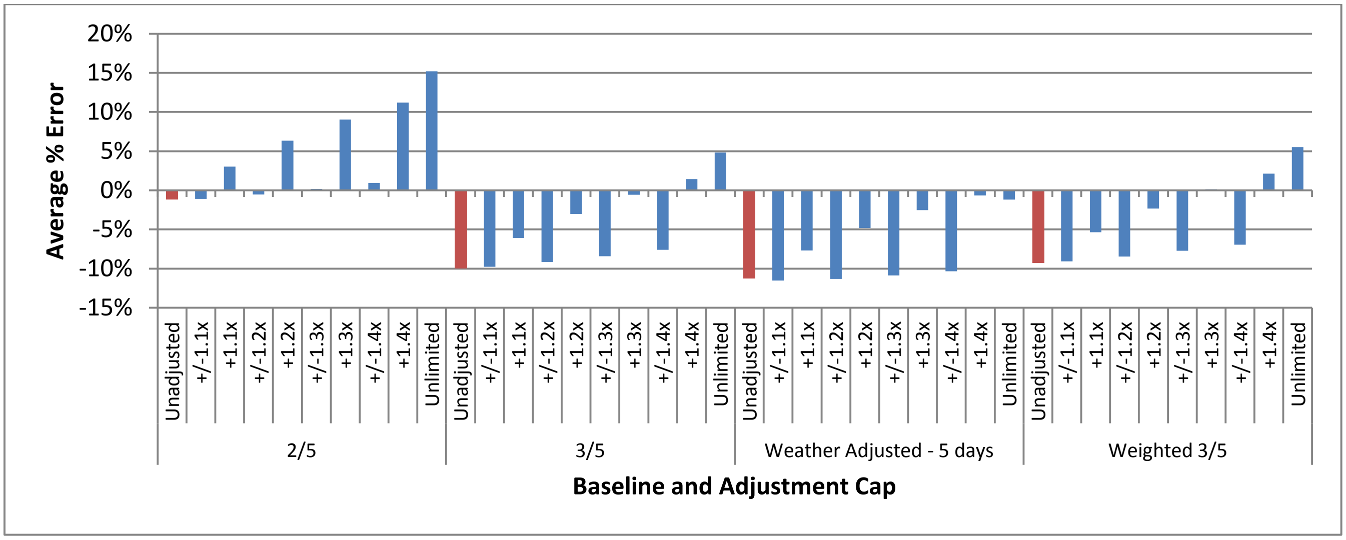


Figure 4-7 shows the effect of applying same-day adjustments on the average percent error, which is a measure of bias. This is the same metric discussed in Section 4.1.1. As before, the unadjusted baseline is shown in red. All baselines experience decreased bias when they are adjusted, except for the 2/5 baseline, which already had very low bias. However, the adjustments that perform well on the average percent error metric (bias) are not the same ones that perform well on the sum of absolute errors metric (precision). To minimize absolute error, the best adjustments are fairly small, around 1.1, whereas to minimize average percent error, the best adjustments are slightly larger, around 1.3. This means that it is difficult to choose a single adjustment that reduces bias and improves precision. In addition, there is some variation across baselines, because the best adjustment for one baseline is not necessarily the best adjustment for another.

Figure 4-7: Average Percent Error (Bias) by Baseline Type and Same-Day Adjustment Cap



Thus, while it is clear that same-day adjustments do make baselines more accurate, FSC is unable to recommend a single adjustment that works for all baselines. More generally, it is not recommended to apply same-day adjustments to PTR baselines because of some of the reasons alluded to previously. The logic behind applying a same-day adjustment only holds if customers do not change their behavior right before an event. However, with PTR, customers are actively encouraged to modify their behavior. Some customers may choose to pre-cool their home, thus increasing their usage before an event and skewing the same-day adjustment upward. Other customers may choose to reduce their usage in anticipation of the event, thus decreasing their usage before the event and skewing the same-day adjustment downward. Both of these behaviors are perfectly legitimate but mean that the main assumption regarding the validity of the same-day adjustment – that customers do not change their behavior before an event – is violated.

4.4 Baseline Accuracy Conclusions

As seen above, no baseline is perfect. All baselines are inaccurate estimates of individual customer's electricity usage on individual days. However, some baselines do perform better than others on most metrics. Table 4-1 lists five unadjusted baseline methods that perform better than most others. For reasons discussed above, we do not recommend using same-day adjustments as they could increase errors for customers who behavior rationally in response to the program incentives. Three of the five baseline methods in Table 4-1 conform to SDG&E's implementation criteria, while two do not. These baseline methods will be used as the basis for assessing payment error under alternative payment rules in Section 5.

Note that the current baseline (3/5) is included in this list. The original research design called for the inclusion of this baseline in the more detailed payment analysis regardless of its accuracy, but the 3/5 baseline is actually accurate enough to be included on its own merit. In fact, there are no other conforming baselines that are able to provide significant improvements over it in terms of baseline accuracy. Given that baseline accuracy cannot be improved dramatically by any of the numerous methods examined, reduction in payment error will need to come from some other strategy. Some options are examined in Section 5.

Table 4-1: Suggested Baselines

Baseline	Type	Conforming
2/5	Day-matching	Yes
3/5	Day-matching	Yes
Weighted 3/5	Day-matching	Yes
Weather Adjusted - 5 days	Weather-matching	No
Regression	Other	No

5 Payment Accuracy Results

This section summarizes an assessment of payment accuracy using the five baseline methods listed in Table 4-1. Payment accuracy is more complicated to understand than impact accuracy. As discussed in Section 3.4.1, PTR payments are asymmetrical – that is, customers are paid for demand reductions (as measured by the baseline) but they are not charged for demand increases (as measured by the baseline). Thus, the direction of the impact error becomes very important. Baselines that consistently underestimate usage tend to have lower payments made in error than those that overestimate usage, but only because downwardly biased baselines pay fewer people overall.

The size of a customer's load impact is also important. Due to PTR's asymmetric payment structure, even if a customer does not produce any load reduction, there is still a high probability that they will get paid. In this case, all payment error is in the customer's favor and all payments are due to baseline measurement error. On the other hand, if a customer provides a relatively large load impact, that customer can be overpaid *or* underpaid, meaning that overpayments and underpayments can cancel each other out and will tend to be more accurate on average across multiple event days.

Payment accuracy results are presented for three scenarios. Each scenario assumes different average load reductions and enrollment levels and is intended to represent what might occur under three different program strategies:

- Scenario 1 is meant to represent a default PTR scenario in which average impacts are small for a large population of customers;
- Scenario 2 is meant to represent an opt in program in which enrollment is much lower than for a default scenario, but average load reductions are higher;
- Scenario 3 represents a more targeted, opt-in program comprised of fewer customers than Scenario 2 but with higher average load reductions. This is achieved by a marketing strategy designed to target and enroll high responders.

The results for each of the scenarios are shown for the current tariff and for the alternate payment mechanisms documented in Section 3.4.2. The performance metrics used are defined in Section 3.4.3.

5.1 Scenario 1: Default Enrollment

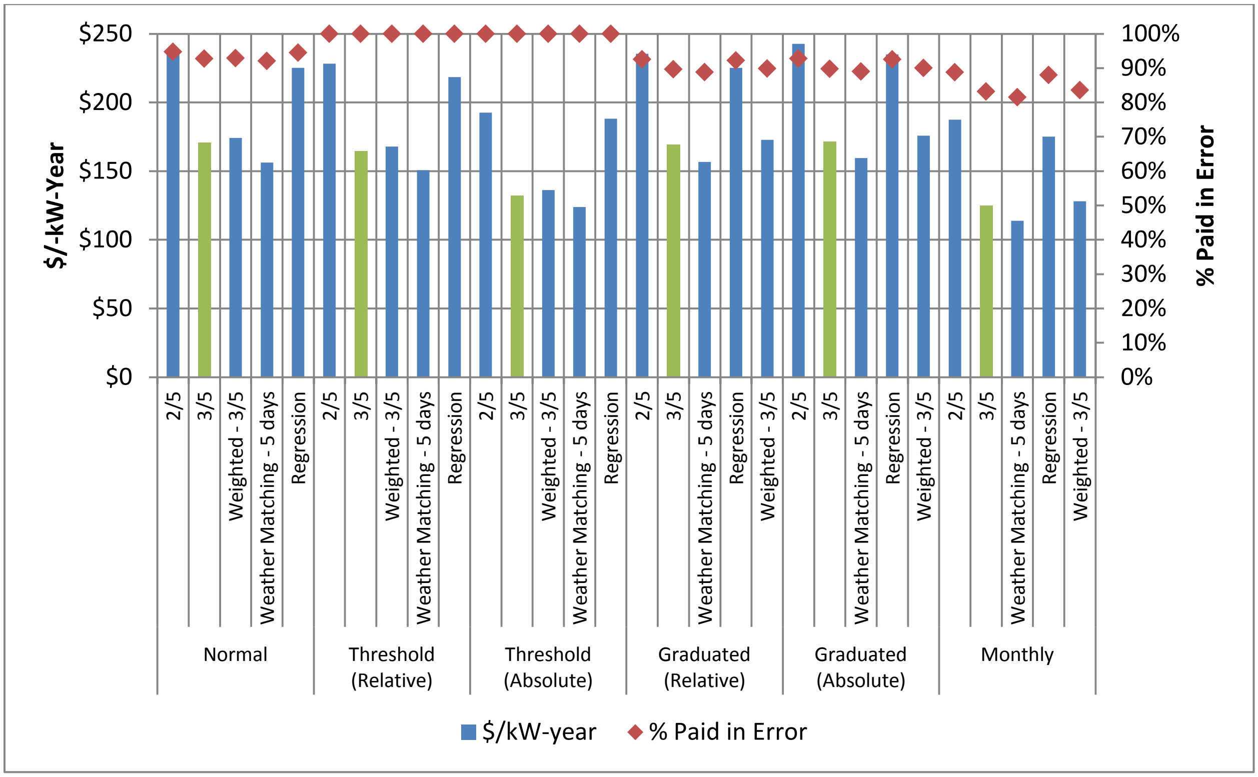
Under the default enrollment scenario, it is assumed that all 1.2 million SDG&E residential customers are defaulted onto PTR and that, on average, they reduce peak period load by 2%.[16] A 2% reduction, when applied to the average customer load on proxy days for SDG&E's population yields an actual reduction of approximately 30 MW.

Figure 5-1 shows an overview of payment accuracy for this default PTR scenario using the five "high performing" baseline methods summarized in Table 4-1 and the six payment rules depicted in Section 3.4.2 (including the current rule). The current baseline (3/5) is shown in green. The figure includes a simple metric for each baseline/payment rule combination, $/kW-yr, which is equal to the total incentives paid divided by the demand reduction obtained (in this example, 30 MW). Dollars per kW-year is a useful metric for comparing the cost of procuring demand reductions through PTR to the cost

[16] The impact evaluation of SDG&E's PTR program showed no statistically significant impact at all for the average default customer, so the assumptions made here are more optimistic than what has been observed so far for the current program.

of procuring capacity from other sources, such as peaking plants or solar arrays.[17] The figure also shows excess payments as a percent of total payments made. A value of 100% means that all of the payments made under that baseline and payment scheme should *not* have been made, whereas a value of 0% means that, overall, the total payment amount accurately reflects what should have been paid. Note that this does not mean that every individual customer was paid correctly, only that overpayments and underpayments to customers exactly offset each other, and the overall total payment was correct.

Figure 5-1: Assessment of Payment Accuracy – Scenario 1 (Default Enrollment)



As Figure 5-1 shows, payment error constitutes between 80% and 100% of total PTR payments across all baseline/payment rule combinations for this scenario involving default-enrollment with small average load reductions. PTR costs per avoided kW range between $110 and $240. Expressed in dollar terms, payment error ranges from $2.8 million to $6.6 million. Under most baseline/payment combinations, more than 90% of customers are paid too much, and the median customer is paid approximately 5 times as much as they should be (the tables underlying these statistics are included in Appendix C).

Figure 5-1 reveals two interesting patterns. First, the 2/5 day-matching baseline and the regression method have some of the highest $/kW-year costs and have relatively high payment errors, even though they are relatively unbiased compared to the other baselines in the figure, which have lower payment errors and costs. The reason for this is that the other baselines are biased downward. In

[17] Obviously, incentive payments are not the only cost of a PTR program. Costs for program start up, program management, marketing and educational campaigns, event notification and other necessary activities would also need to be included in order to make an "apples to apples" comparison with alternative resources. Nevertheless, $/kW-yr is a useful, simple metric for comparing the relative costs of different baseline/payment rule combinations.

general, this causes them to underpay all customers, even those that provide demand reductions. In a scenario where demand reductions are very small, as they are in this case, PTR's asymmetrical payment structure causes baselines that are biased downward to generate lower payments made in error – but only because they are underpaying most customers. The current 3/5 baseline, shown in green, performs relatively well because it has a downward bias of about 10%, as shown previously in Figure 4-1.

Second, the two worst performing payment schemes in terms of payment error – the relative and absolute thresholds – actually have among the lowest values for $/kW-yr. The reason for this is that, as their name implies, these schemes set a demand reduction threshold below which payments are not made. In this scenario, the 2% demand reduction is not enough to exceed the payment threshold for many customers, causing the threshold payment schemes to not pay many customers who do actually reduce their load. However, if the baseline error is large enough, the estimated reduction will still exceed the threshold and some customers will get paid. Although under these schemes, all payments are made in error, overall payments are minimized; few customers get paid anything, since their estimated reductions are below the threshold.

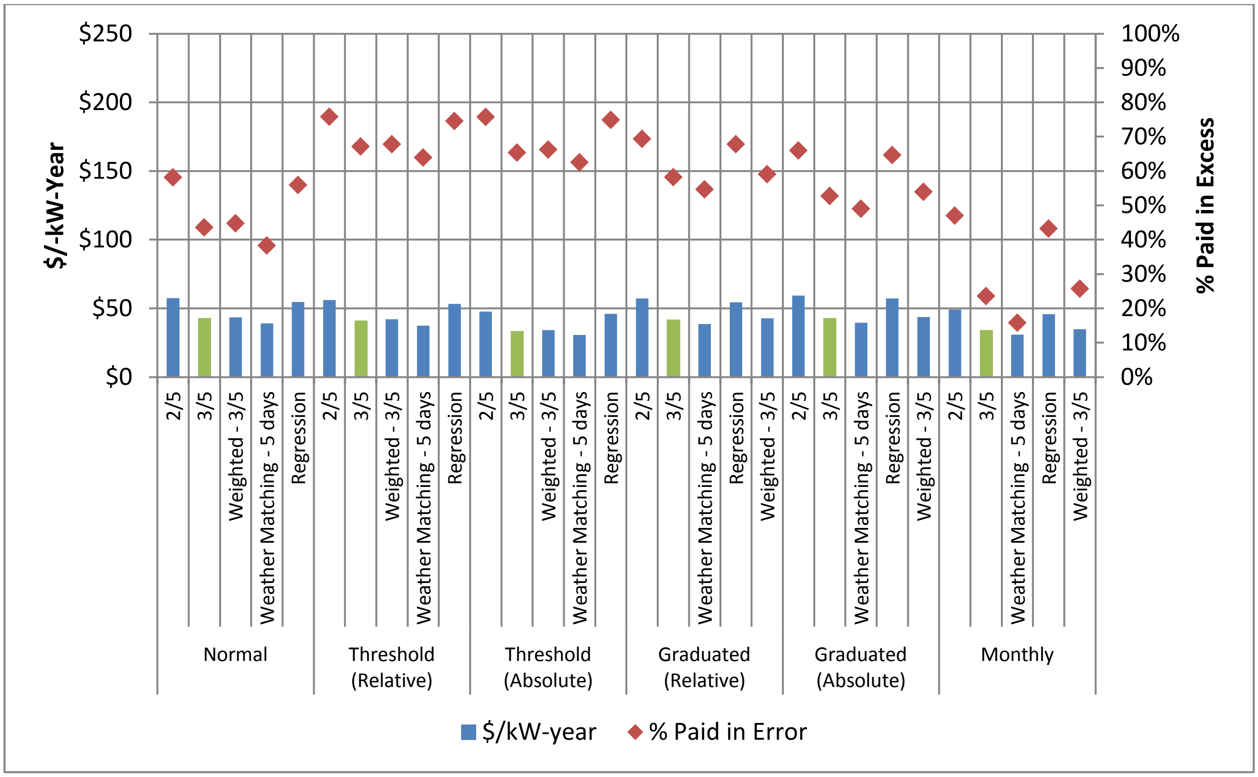
Of the six payment rules depicted in Figure 5-1, the best performing rule under this scenario is the monthly settlement method, which uses the total load reduction for all event days in a month. This rule, shown on the right of the figure, has some of the lowest overall percent payment error, although payment error is still quite high, ranging from 80% to 90%. This approach also has among the lowest $/kW-yr values. When this rule is combined with the current 3/5 baseline method, the avoided cost metric is around $125/kW-yr.

5.2 Scenario 2: Opt-in Enrollment

Under this opt-in enrollment scenario, it is assumed that 20%, or 240,000 of SDG&E's residential customers, participate in PTR. It is also assumed that customers on average, reduce load by 10%. The 10% impact was chosen for illustrative purposes only and is not intended as a forecast of the kind of load impacts that could be expected under an opt-in enrollment program format. This load reduction estimate is higher than the 2% assumed in the previous scenario because customers who opt in to the program are much more likely to do something when a PTR event is called, partly because they are more likely to be motivated to respond and partly because personalized event notification using email or text messages will be much more common in an opt-in deployment than under default enrollment. When multiplying the 10% impact across 240,000 customers, this scenario again yields approximately 30 MW of actual reduction – the same reduction as in Scenario 1. Again, the 30 MW of actual reduction is not a forecast of the likely load impacts from opt-in PTR; this number was chosen for illustrative purposes only.

Figure 5-2 shows the cost per kW-year and the percent of customer payments made in excess for each baseline/payment rule combination.

Figure 5-2: Payment Accuracy – Scenario 2 (Opt-in Enrollment, No Targeting)



The contrast between Figure 5-1 and Figure 5-2 is striking. The cost per kW-year is markedly lower, as is the percent of the customer payments made in error. Expressed in dollar terms, the payment error also decreased dramatically. On average, assuming five events per year, it is about $750,000; for the monthly settlement method, aggregate payment error only equals $150,000. As before, the current 3/5 baseline performs relatively well. Although many customers are still getting overpaid for their reductions, both the proportion of customers getting overpaid and the total amount of overpayment is much lower under this scenario than for Scenairo 1.

The improvements under this scenario can be attributed to two factors. First, because the actual impact is assumed to be 10% – five times larger than the impact of 2% assumed under Scenario 1 – the average impact error is five times smaller than under Scenario 1.[18] This translates into considerably smaller payment errors. Second, because the actual and estimated impacts are larger, the payment asymmetry inherent in PTR is reduced since some customers are underpaid while others are overpaid. Recall that when impacts are very small, the only payment error that can occur is in the customer's favor; however, when impacts are larger, payment errors occur in both directions and can start to cancel out, leading to lower overall excess payment.

It is important to note that, when comparing the $/kW-yr values for Scenarios 1 and 2, this metric only factors in load reduction incentive costs, not overall program costs. Program costs are likely to be higher for an opt-in program than for a default program because of marketing and recruitment

---

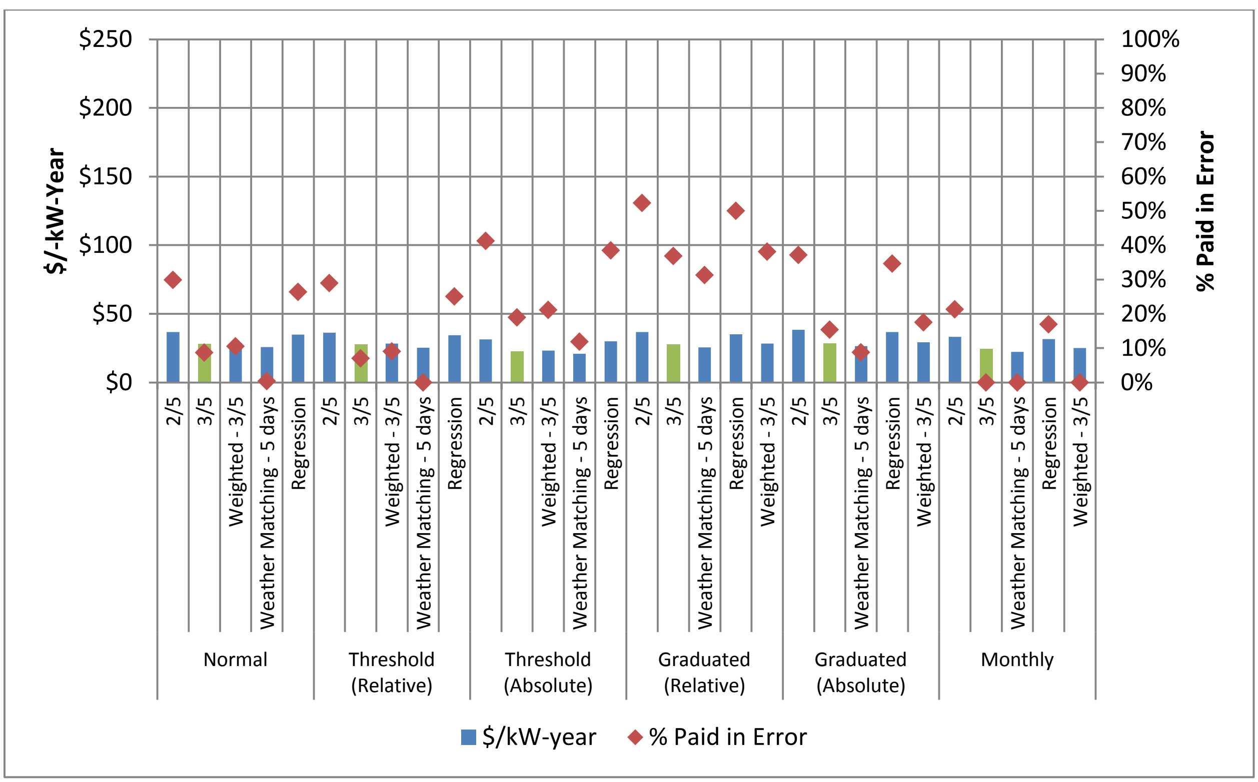
[18] As shown in Section 3, this is because baseline error and impact error are directly related: $\frac{baseline\ error\ (\%)}{impact\ size\ (\%)} = impact\ error\ (\%)$. If the relative impact increases five-fold, the denominator in this equation increases fivefold, meaning that the impact error becomes one fifth its previous size.

costs. However, the reduction in excess payments between the two scenarios should more than offset marketing and recruitment costs, especially over multiple years. Excess payments are made every year the program exists (as long as events are called), while marketing and recruitment costs fall significantly once a program reaches a steady-state enrollment level. Under default enrollment with the assumptions made for Scenario 1, excess payments amount to $2.8 to $6.6 million every year, while in the opt-in scenario, excess payments are as low as $150,000 per year and only equal $750,000 per year under the worst case baseline/payment rule combination. Given the no-risk nature of PTR, acceptance rates should be reasonably high and marketing costs relatively low. It is hard to imagine that marketing costs for opt-in enrollment would be so high as to offset the benefit of avoiding excess payments of many millions of dollars each year associated with the default program.

5.3 Scenario 3: Targeted Enrollment

The third scenario assumes that SDG&E targets and recruits high responders. From prior PTR and CPP pilots, it is known that high responders are more likely to have central air conditioning. Moreover, through analysis of interval data, customers with air conditioning can be identified with high confidence and marketing activities can focus on this segment. Targeting can be improved and refined over time based on analysis of early adopters. For the analysis here, we have assumed that SDG&E would be able to enroll 120,000 high responders into the program, and that the average reduction per participant would be 20%. As with the other two scenarios, this scenario yields approximately 30 MW of load reduction. Again, this figure is used for illustrative purposes only and is not to meant to be used as a forecast for the load impact to be expected from PTR with targeted enrollment. Figure 5-3 summarizes the results for this scenario based on the above assumptions.

Figure 5-3: Payment Accuracy – Scenario 3 (Opt-in Enrollment with Targeting)



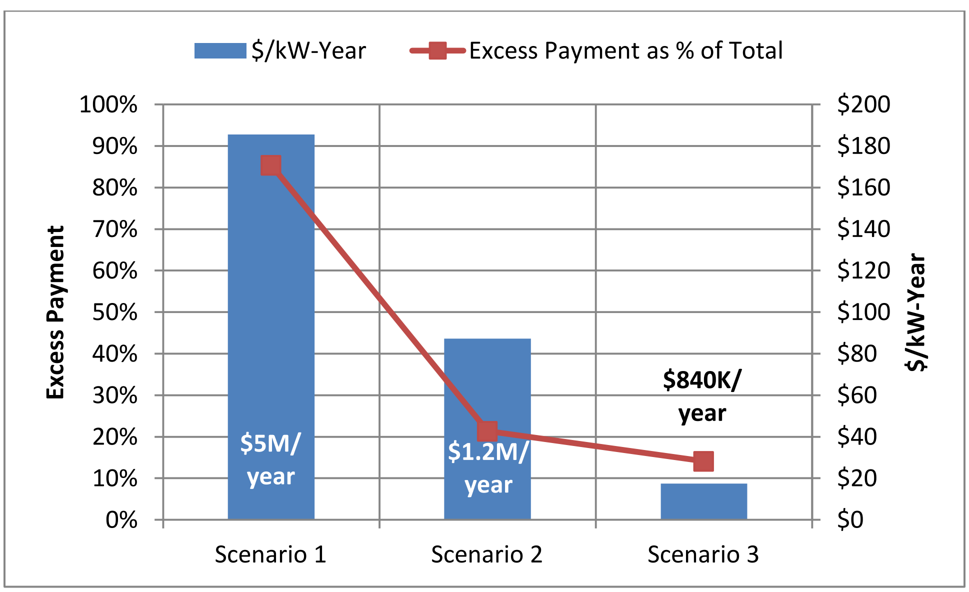
Compared with the previous two scenarios, Scenario 3 shows very significant reductions in both the percentage of customer payments made in error and the cost per kW-year. The cost per kW-yr ranges from $21 to $38. Several baseline/payment combinations have no excess payments – that is, overpayments and underpayments to individual customers on individual event days completely cancel out. In fact, for some combinations – notably, several of the monthly averaging baseline methods – underpayments more than offset overpayments, and the average customer is paid less than they should be, considering the load reduction they provide. Most baseline and payment combinations still overpay more customers than they underpay, but payment errors are generally much smaller.

5.4 Payment Accuracy Conclusions

From the analysis presented in this section, it is clear that the best way to minimize excess payments is to change the structure of the program, rather than refine the baseline or change the payment rules. As each successive scenario shows, the best way to avoid aggregate overpayments is to target customers that provide larger load impacts. This also requires reducing enrollment of customers that do not reduce demand but still receive payment due to baseline error and payment asymmetry. Larger customer impacts reduce the effect of PTR's payment asymmetry and minimize the program's cost per kW-yr. To achieve these benefits, FSC suggests changing program enrollment from a default basis to an opt-in basis and, if possible, targeting customers who can provide larger load reductions.

Figure 5-4 shows a graphical overview of the implications for each of the scenarios using the current baseline. Scenario 1 – default PTR – has costs of $5 million per year, with high error rates and high rebate payments per MW of demand reduction delivered. Scenario 2 – opt-in PTR – has costs of $1.2 million per year, with lower error rates and lower costs per MW. Scenario 3 – targeted opt-in PTR – has costs of $840,000 per year with the lowest errors of all. All scenarios provide approximately 30 MW of reduction, but for vastly different price tags. The policy implications are clear – rather than changing baseline methods and payment rules, the best way to improve PTR cost effectiveness is to shift away from default to opt-in enrollment, targeting customers that deliver large average demand reductions.

Figure 5-4: Scenario Overview



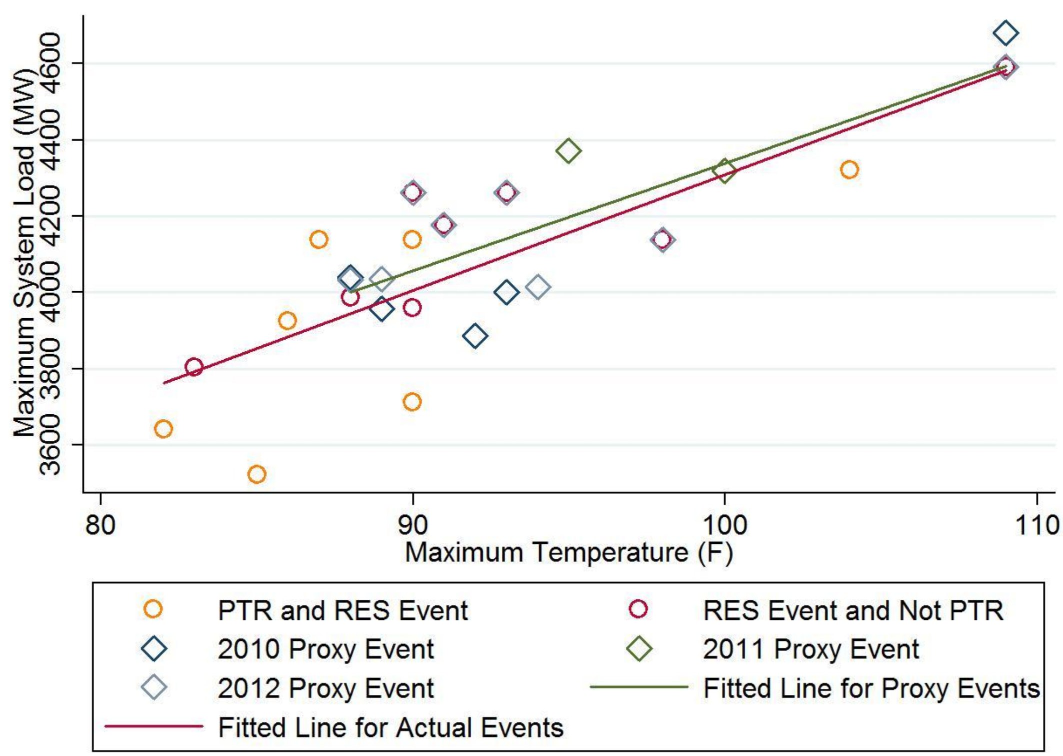
Appendix A Proxy Event Selection

Figure A-1 shows a graphical comparison between actual event days and the proxy event days that were used to test baseline accuracy. The accuracy of the simulation exercise documented in this report hinges on the selection of proxy event days that are as similar to actual PTR event days as possible.

Most actual PTR events in 2012 were called early in the summer. A conscious choice was made not to call more events later in the summer, even though weather and system conditions were appropriate to do so. In order to understand the full range of weather and system conditions under which PTR events were called, other residential DR events were also included in the library of actual events used to generate proxy events. This increases the number of data points available for comparison and ensures that proxy events are as close to actual event conditions as possible, even though a choice was made to call fewer actual PTR events than might have otherwise been called.

Figure A-1, which compares conditions on actual event days to conditions on proxy event days, shows 2012 PTR DR events, 2012 non-PTR DR events, and proxy events called in 2010, 2011, and 2012. Because the algorithm used to generate proxy events for the evaluation selected several non-PTR DR event days as proxy events, it is clear that proxy event conditions match actual event conditions. The green line shows a trend line between load and temperature for proxy events; the red line shows the same thing for actual events. Because the two lines match each other quite well, it is clear that proxy events and actual events are very similar.

Figure A-1: Actual Event Days Compared to Proxy Event Days



Appendix B Detailed Baseline Accuracy Tables

B.1 Comparisons Across Baselines

This section shows baseline accuracy statistics for the average event day for all baselines tested in the study.

Table B-1: Baseline Accuracy of Unadjusted Baselines during the 11 AM – 6 PM Event Window (Average Event Day)

Baseline	Program Level Results			Customer-Level Results					
	Sum of Errors	Errors Squared	Average Errors	MPE (Percentiles)			MAPE (Percentiles)		
				10	50	90	10	50	90
Regression	88,760	825,256	-1%	-24%	0%	24%	2%	11%	34%
Weather Matching - 4 days	89,303	917,035	-10%	-30%	-6%	17%	2%	12%	35%
Weather Matching - 5 days	89,315	913,490	-11%	-31%	-7%	16%	2%	13%	35%
Weather Matching - 3 days	90,532	947,241	-9%	-29%	-6%	19%	2%	12%	34%
Weighted - 3/5	90,878	1,005,918	-9%	-30%	-1%	27%	2%	14%	39%
3/5	91,407	1,011,812	-10%	-31%	-1%	28%	2%	14%	40%
3/5 - +5%	92,334	1,003,418	-5%	-27%	4%	34%	3%	15%	42%
2/5	93,063	1,001,575	-1%	-23%	6%	43%	3%	15%	45%
4/5	93,458	1,081,942	-17%	-38%	-8%	16%	2%	14%	41%
Weighted - 3/3	94,350	1,148,255	-20%	-40%	-12%	8%	2%	15%	42%
5/10	94,727	1,019,174	-7%	-28%	4%	37%	3%	16%	44%
3/3	96,544	1,192,021	-22%	-42%	-13%	8%	2%	16%	43%
5/5	98,069	1,190,848	-24%	-45%	-14%	8%	2%	17%	46%
Weather Adjusted - 3/5	98,404	1,015,160	8%	-17%	17%	53%	5%	21%	54%
3/10	98,484	1,034,334	5%	-17%	13%	58%	4%	18%	58%
10/20	99,920	1,090,576	-9%	-33%	3%	40%	3%	17%	48%
10/10	102,769	1,270,917	-26%	-47%	-14%	8%	3%	18%	48%
5/20	105,006	1,077,884	7%	-18%	17%	69%	4%	22%	69%
Weather - sumcdh	107,134	1,164,429	-18%	-38%	-12%	13%	3%	16%	40%
Weather - cdd	107,638	1,174,790	-14%	-35%	-10%	16%	3%	15%	39%
Weather - maxtemp	110,474	1,250,605	-21%	-42%	-14%	12%	3%	17%	44%
THI	118,571	1,262,034	27%	1%	33%	115%	6%	33%	115%

*Payments are calculated assuming impacts of 0%. Total payments assume there are 1.2 million customers and 15 event days.

All results are given for the average event day.

Table B-2: Baseline Accuracy of +1.2 Adjusted Baselines during the 11 AM – 6 PM Event Window (Average Event Day)

Baseline	Program Level Results			Customer-Level Results					
	Sum of Errors	Errors Squared	Average Errors	MPE (Percentiles)			MAPE (Percentiles)		
				10	50	90	10	50	90
Regression	88,760	825,256	-1%	-24%	0%	24%	2%	11%	34%
Weather Matching - 4 days	86,907	852,643	-4%	-24%	0%	24%	2%	11%	34%
Weather Matching - 5 days	86,121	834,465	-5%	-25%	-1%	23%	2%	12%	33%
Weather Matching - 3 days	88,966	895,940	-2%	-23%	1%	26%	2%	11%	36%
Weighted - 3/5	90,865	964,116	-2%	-24%	6%	35%	3%	15%	40%
3/5	91,013	961,330	-3%	-25%	6%	35%	3%	15%	41%
3/5 - +5%	93,623	981,572	1%	-22%	10%	41%	4%	17%	44%
2/5	96,623	1,016,624	6%	-16%	14%	52%	4%	18%	52%
4/5	89,968	983,850	-11%	-32%	-1%	24%	2%	14%	39%
Weighted - 3/3	90,711	1,053,016	-14%	-35%	-5%	15%	2%	13%	38%
5/10	94,005	962,264	0%	-23%	10%	45%	4%	18%	47%
3/3	92,419	1,086,539	-15%	-36%	-6%	15%	2%	13%	39%
5/5	91,964	1,054,049	-18%	-39%	-7%	14%	2%	14%	41%
Weather Adjusted - 3/5	102,548	1,062,980	14%	-12%	22%	59%	5%	24%	60%
3/10	102,443	1,053,121	12%	-11%	20%	66%	4%	22%	66%
10/20	97,913	1,007,802	-2%	-28%	10%	48%	4%	19%	51%
10/10	95,687	1,109,517	-20%	-43%	-8%	15%	2%	15%	45%
5/20	108,130	1,084,395	15%	-11%	24%	78%	5%	26%	78%
Weather - sumcdh	100,545	1,015,350	-12%	-33%	-5%	18%	2%	13%	37%
Weather - cdd	102,815	1,058,535	-7%	-30%	-3%	23%	2%	13%	37%
Weather - maxtemp	103,244	1,083,973	-15%	-37%	-6%	18%	2%	14%	41%
THI	128,440	1,401,352	35%	8%	41%	124%	9%	41%	124%

*Payments are calculated assuming impacts of 0%. Total payments assume there are 1.2 million customers and 15 event days.

All results are given for the average event day.

Table B-3: Baseline Accuracy of +/-1.2 Adjusted Baselines during the 11 AM – 6 PM Event Window (Average Event Day)

Baseline	Program Level Results			Customer-Level Results					
	Sum of Errors	Errors Squared	Average Errors	MPE (Percentiles)			MAPE (Percentiles)		
				10	50	90	10	50	90
Regression	88,760	825,256	-1%	-24%	0%	24%	2%	11%	34%
Weather Matching - 4 days	86,544	850,124	-10%	-30%	-7%	15%	2%	12%	34%
Weather Matching - 5 days	86,108	841,064	-11%	-31%	-7%	14%	2%	12%	34%
Weather Matching - 3 days	88,150	880,813	-9%	-28%	-6%	16%	2%	12%	34%
Weighted - 3/5	89,818	957,113	-8%	-29%	-1%	25%	2%	13%	38%
3/5	90,027	957,680	-9%	-30%	-1%	25%	2%	13%	38%
3/5 - +5%	90,657	957,176	-6%	-27%	2%	29%	2%	14%	39%
2/5	92,848	966,032	-1%	-22%	6%	39%	3%	15%	43%
4/5	90,965	1,011,759	-16%	-37%	-7%	15%	2%	13%	40%
Weighted - 3/3	92,681	1,087,640	-19%	-39%	-11%	8%	2%	15%	40%
5/10	91,266	937,949	-7%	-28%	2%	33%	2%	15%	40%
3/3	94,376	1,123,734	-20%	-40%	-13%	8%	2%	15%	42%
5/5	94,478	1,105,408	-23%	-44%	-13%	7%	2%	16%	45%
Weather Adjusted - 3/5	92,755	944,185	3%	-20%	10%	43%	4%	16%	45%
3/10	95,782	955,719	4%	-17%	11%	52%	3%	17%	52%
10/20	94,981	992,925	-9%	-33%	2%	36%	3%	16%	44%
10/10	98,006	1,163,432	-25%	-46%	-14%	7%	3%	16%	47%
5/20	99,411	965,258	6%	-18%	14%	62%	4%	19%	62%
Weather - sumcdh	100,571	1,047,664	-18%	-38%	-12%	9%	2%	15%	40%
Weather - cdd	101,813	1,060,547	-15%	-35%	-10%	13%	2%	14%	38%
Weather - maxtemp	103,676	1,127,907	-21%	-42%	-13%	8%	3%	16%	44%
THI	112,682	1,114,885	24%	0%	30%	104%	5%	30%	104%

*Payments are calculated assuming impacts of 0%. Total payments assume there are 1.2 million customers and 15 event days.

All results are given for the average event day.

B.2 Comparisons Across Dates

This section shows the baseline accuracy for each of the proxy event days for each of the five baselines included in the payment analysis.

Table B-4: Daily Baseline Accuracy of the Unadjusted 3/5 Baseline during the 11 AM – 6 PM Event Window

Proxy Event Date	Average Customer Maximum Temperature	Program Level Results			Customer-Level Results					
		Sum of Errors	Errors Squared	Average Error	Percent Error (Percentiles)			Absolute Percent Error (Percentiles)		
					10	50	90	10	50	90
8/18/2010	87	6,141	65,571	-18%	-51%	-2%	83%	4%	26%	85%
8/19/2010	85	5,812	61,078	-16%	-50%	-2%	82%	4%	25%	84%
8/23/2010	86	4,766	42,337	-4%	-37%	3%	94%	2%	22%	95%
8/26/2010	83	4,419	34,845	7%	-22%	9%	105%	1%	19%	105%
9/27/2010	101	11,849	234,897	-47%	-78%	-17%	62%	6%	38%	87%
9/6/2011	96	6,661	72,885	-5%	-47%	3%	98%	4%	28%	100%
9/7/2011	90	7,281	84,256	-19%	-54%	-3%	82%	4%	27%	86%
8/13/2012	86	6,014	62,837	-9%	-45%	-1%	88%	3%	22%	92%
8/16/2012	84	5,373	46,130	8%	-30%	10%	120%	3%	23%	120%
8/17/2012	87	6,552	70,361	-7%	-44%	1%	95%	4%	24%	96%
8/31/2012	83	4,615	33,773	2%	-33%	5%	91%	3%	21%	94%
9/4/2012	87	4,681	35,851	7%	-28%	10%	109%	3%	21%	109%
9/14/2012	97	6,805	71,266	-13%	-45%	2%	90%	4%	26%	92%
10/1/2012	87	5,072	46,705	-11%	-44%	1%	74%	4%	23%	79%
10/2/2012	91	5,364	49,017	-13%	-46%	1%	83%	4%	25%	86%
Average Day	**89**	**6,094**	**67,454**	**-10%**	**-31%**	**-1%**	**28%**	**2%**	**14%**	**40%**
Total	**-**	**91,407**	**1,011,812**	**-**	**-**	**-**	**-**	**-**	**-**	**-**

Table B-5: Daily Baseline Accuracy of the Unadjusted 2/5 Baseline during the 11 AM – 6 PM Event Window

Proxy Event Date	Average Customer Maximum Temperature	Program Level Results			Customer-Level Results					
		Sum of Errors	Errors Squared	Average Error	Percent Error (Percentiles)			Absolute Percent Error (Percentiles)		
					10	50	90	10	50	90
8/18/2010	87	5,780	56,804	-8%	-43%	3%	107%	3%	26%	107%
8/19/2010	85	5,533	53,990	-6%	-42%	3%	103%	3%	25%	103%
8/23/2010	86	4,990	43,709	7%	-29%	11%	121%	2%	24%	121%
8/26/2010	83	5,040	42,455	17%	-15%	16%	137%	1%	23%	137%
9/27/2010	101	11,581	222,197	-42%	-76%	-12%	81%	5%	39%	90%
9/6/2011	96	6,851	75,297	4%	-40%	9%	126%	4%	29%	126%
9/7/2011	90	6,982	74,799	-11%	-48%	2%	104%	4%	28%	104%
8/13/2012	86	6,093	62,770	-1%	-38%	5%	112%	3%	22%	112%
8/16/2012	84	6,021	54,475	18%	-23%	18%	151%	3%	27%	151%
8/17/2012	87	6,640	70,184	2%	-36%	8%	119%	4%	25%	119%
8/31/2012	83	4,946	38,310	12%	-24%	12%	116%	3%	23%	116%
9/4/2012	87	5,324	44,301	17%	-21%	17%	138%	3%	25%	138%
9/14/2012	97	6,770	68,053	-5%	-40%	7%	118%	4%	27%	118%
10/1/2012	87	5,145	46,224	-3%	-38%	7%	99%	4%	24%	99%
10/2/2012	91	5,369	48,009	-4%	-39%	6%	108%	4%	26%	108%
Average Day	**89**	**6,204**	**66,772**	**-1%**	**-23%**	**6%**	**43%**	**3%**	**15%**	**45%**
Total	**-**	**93,063**	**1,001,575**	**-**	**-**	**-**	**-**	**-**	**-**	**-**

Table B-6: Daily Baseline Accuracy of the Unadjusted Weighted 3/5 Baseline during the 11 AM – 6 PM Event Window

Proxy Event Date	Average Customer Maximum Temperature	Program Level Results			Customer-Level Results					
		Sum of Errors	Errors Squared	Average Error	Percent Error (Percentiles)			Absolute Percent Error (Percentiles)		
					10	50	90	10	50	90
8/18/2010	87	5,774	58,267	-14%	-46%	-1%	85%	3%	25%	87%
8/19/2010	85	5,484	55,396	-13%	-45%	-1%	82%	3%	24%	85%
8/23/2010	86	4,633	39,840	-1%	-34%	4%	104%	2%	21%	104%
8/26/2010	83	4,350	34,124	9%	-20%	9%	107%	1%	19%	107%
9/27/2010	101	11,747	231,224	-46%	-78%	-17%	60%	6%	37%	87%
9/6/2011	96	6,784	75,754	-8%	-49%	2%	99%	4%	28%	99%
9/7/2011	90	7,538	90,737	-21%	-58%	-4%	79%	4%	28%	85%
8/13/2012	86	5,938	62,500	-7%	-43%	0%	87%	3%	21%	91%
8/16/2012	84	5,302	45,092	9%	-28%	11%	117%	3%	23%	117%
8/17/2012	87	6,521	68,714	-6%	-42%	1%	93%	4%	24%	93%
8/31/2012	83	4,594	33,833	4%	-32%	5%	93%	3%	20%	94%
9/4/2012	87	4,758	37,301	9%	-26%	10%	108%	3%	21%	108%
9/14/2012	97	6,809	71,958	-14%	-47%	1%	90%	4%	25%	91%
10/1/2012	87	5,175	49,101	-13%	-45%	0%	75%	3%	23%	80%
10/2/2012	91	5,473	52,076	-14%	-48%	0%	81%	4%	25%	86%
Average Day	**89**	**6,059**	**67,061**	**-9%**	**-30%**	**-1%**	**27%**	**2%**	**14%**	**39%**
Total	**-**	**90,878**	**1,005,918**	**-**	**-**	**-**	**-**	**-**	**-**	**-**

Table B-7: Daily Baseline Accuracy of the Unadjusted Weather Matching 5 Day Baseline during the 11 AM – 6 PM Event Window

Proxy Event Date	Average Customer Maximum Temperature	Program Level Results			Customer-Level Results					
		Sum of Errors	Errors Squared	Average Error	Percent Error (Percentiles)			Absolute Percent Error (Percentiles)		
					10	50	90	10	50	90
8/18/2010	87	5,386	52,569	-13%	-48%	-3%	76%	2%	23%	82%
8/19/2010	85	5,520	51,751	-13%	-48%	-5%	82%	4%	24%	84%
8/23/2010	86	4,873	42,474	-3%	-40%	0%	96%	2%	23%	98%
8/26/2010	83	4,826	44,513	-10%	-41%	0%	74%	2%	22%	75%
9/27/2010	101	8,008	112,178	-23%	-57%	-8%	78%	2%	28%	87%
9/6/2011	96	6,752	69,816	-2%	-51%	2%	118%	5%	30%	118%
9/7/2011	90	7,023	80,161	-15%	-56%	-5%	92%	4%	27%	94%
8/13/2012	86	7,045	82,634	-24%	-57%	-11%	57%	5%	26%	77%
8/16/2012	84	5,222	46,557	-11%	-44%	-3%	76%	4%	22%	81%
8/17/2012	87	7,321	86,574	-25%	-56%	-12%	58%	5%	28%	77%
8/31/2012	83	4,923	41,583	-7%	-41%	0%	82%	3%	22%	86%
9/4/2012	87	4,658	34,035	1%	-35%	3%	104%	3%	22%	104%
9/14/2012	97	7,104	79,352	-16%	-52%	-5%	82%	4%	26%	90%
10/1/2012	87	5,353	45,778	3%	-39%	4%	108%	4%	25%	108%
10/2/2012	91	5,300	43,514	2%	-41%	4%	103%	4%	24%	103%
Average Day	**89**	**5,954**	**60,899**	**-11%**	**-31%**	**-7%**	**16%**	**2%**	**13%**	**35%**
Total	**-**	**89,315**	**913,490**	**-**	**-**	**-**	**-**	**-**	**-**	**-**

Table B-8: Daily Baseline Accuracy of the Unadjusted Regression Baseline during the 11 AM – 6 PM Event Window

Proxy Event Date	Average Customer Maximum Temperature	Program Level Results			Customer-Level Results					
		Sum of Errors	Errors Squared	Average Error	Percent Error (Percentiles)			Absolute Percent Error (Percentiles)		
					10	50	90	10	50	90
8/18/2010	87	5,147	40,311	0%	-35%	3%	115%	4%	24%	118%
8/19/2010	85	4,961	36,381	-5%	-37%	1%	98%	4%	24%	102%
8/23/2010	86	4,912	35,538	4%	-34%	6%	119%	4%	24%	123%
8/26/2010	83	5,433	51,840	-13%	-45%	-2%	77%	4%	26%	88%
9/27/2010	101	7,799	86,301	-1%	-51%	3%	144%	5%	31%	159%
9/6/2011	96	7,613	82,381	11%	-53%	11%	171%	6%	37%	182%
9/7/2011	90	6,747	70,108	-7%	-51%	-1%	122%	5%	30%	127%
8/13/2012	86	5,965	54,231	-6%	-42%	0%	101%	4%	24%	105%
8/16/2012	84	5,063	39,285	-1%	-37%	4%	107%	4%	23%	107%
8/17/2012	87	5,942	53,645	0%	-44%	3%	114%	4%	25%	122%
8/31/2012	83	5,098	40,330	2%	-38%	3%	111%	4%	25%	117%
9/4/2012	87	4,401	29,131	-2%	-35%	3%	95%	4%	22%	98%
9/14/2012	97	8,687	117,710	10%	-55%	8%	158%	6%	35%	177%
10/1/2012	87	5,396	42,910	-3%	-45%	1%	114%	5%	28%	129%
10/2/2012	91	5,597	45,153	-3%	-46%	1%	119%	5%	30%	125%
Average Day	**89**	**5,917**	**55,017**	**-1%**	**-24%**	**0%**	**24%**	**2%**	**11%**	**34%**
Total	**-**	**88,760**	**825,256**	**-**	**-**	**-**	**-**	**-**	**-**	**-**

Appendix C Detailed Payment Accuracy Tables

This section shows detailed results of the payment analysis.

Table C-1: Payment Accuracy Tables for Unadjusted Baselines during the 11AM – 6PM Event Window - Scenario 1: Default Enrollment (2% Impacts, 1.2 Million Customers, 5 Events)

Payment Method	Baseline	Total Payment				$/kW-year	Payment Accuracy			Individual Customer Payment Error Pctiles		
		$Actual	$ Estimated	$ Error	% of Payment in Error		Under paid	Paid Correctly	Over-paid	10	50	90
Normal	2/5	$368,850	$7,040,250	$6,671,400	95%	$235.91	1%	4%	96%	125%	850%	3750%
	3/5		$5,095,800	$4,726,950	93%	$170.75	2%	6%	93%	15%	550%	2700%
	Weighted - 3/5		$5,200,800	$4,831,950	93%	$174.27	2%	5%	93%	25%	557%	2900%
	CAP - 5 days - no forward		$4,663,050	$4,294,200	92%	$156.25	1%	8%	91%	43%	542%	2600%
	Regression		$6,718,650	$6,349,800	95%	$225.13	0%	6%	94%	223%	1000%	3800%
Threshold (Relative)	2/5	-	$6,812,850	$6,812,850	100%	$228.29	0%	4%	96%	-	-	-
	3/5		$4,913,100	$4,913,100	100%	$164.63	0%	6%	94%	-	-	-
	Weighted - 3/5		$5,010,300	$5,010,300	100%	$167.89	0%	5%	95%	-	-	-
	CAP - 5 days - no forward		$4,497,000	$4,497,000	100%	$150.69	0%	8%	92%	-	-	-
	Regression		$6,516,000	$6,516,000	100%	$218.34	0%	6%	94%	-	-	-
Threshold (Absolute)	2/5	-	$5,746,950	$5,746,950	100%	$192.57	0%	22%	78%	-	-	-
	3/5		$3,946,050	$3,946,050	100%	$132.23	0%	31%	69%	-	-	-
	Weighted - 3/5		$4,063,650	$4,063,650	100%	$136.17	0%	30%	70%	-	-	-
	CAP - 5 days - no forward		$3,694,800	$3,694,800	100%	$123.81	0%	37%	63%	-	-	-
	Regression		$5,614,350	$5,614,350	100%	$188.13	0%	30%	70%	-	-	-
Graduated (Relative)	2/5	$522,094	$7,026,570	$6,504,476	93%	$235.45	1%	0%	99%	266%	1238%	4101%
	3/5		$5,055,172	$4,533,078	90%	$169.39	3%	0%	96%	120%	840%	3015%
	CAP - 5 days - no forward		$4,679,188	$4,157,094	89%	$156.79	4%	0%	96%	92%	673%	2538%
	Regression		$6,723,504	$6,201,410	92%	$225.29	1%	0%	99%	222%	1004%	3144%
	Weighted - 3/5		$5,158,846	$4,636,752	90%	$172.86	3%	0%	97%	134%	865%	2958%
Graduated (Absolute)	2/5	$522,094	$7,243,170	$6,721,076	93%	$242.71	1%	0%	99%	267%	1113%	3616%
	3/5		$5,114,088	$4,591,994	90%	$171.36	3%	0%	97%	124%	741%	2568%
	CAP - 5 days - no forward		$4,758,514	$4,236,420	89%	$159.45	4%	0%	95%	84%	580%	2136%
	Regression		$7,013,058	$6,490,964	93%	$235.00	1%	0%	99%	209%	886%	2700%
	Weighted - 3/5		$5,246,390	$4,724,296	90%	$175.80	3%	1%	97%	133%	761%	2591%
Monthly	2/5	$626,700	$5,594,850	$4,968,150	89%	$187.47	6%	4%	89%	-40%	533%	2258%
	3/5		$3,730,500	$3,103,800	83%	$125.00	13%	7%	81%	-100%	300%	1500%
	CAP - 5 days - no forward		$3,394,800	$2,768,100	82%	$113.75	10%	9%	81%	-92%	267%	1400%
	Regression		$5,223,150	$4,596,450	88%	$175.02	4%	7%	89%	20%	514%	2000%
	Weighted - 3/5		$3,819,000	$3,192,300	84%	$127.97	12%	6%	82%	-100%	300%	1542%

Table C-2: Payment Accuracy Tables for Unadjusted Baselines during the 11AM – 6PM Event Window - Scenario 2: Opt-in Enrollment (10% Impacts, 240,000 Customers, and 5 Events)

Payment Method	Baseline	Total Payment				$/kW-year	Payment Accuracy			Individual Customer Payment Error Pctiles		
		$ Actual	$ Estimated	$ Error	% of Payment in Error		Under-paid	Paid Correctly	Over-paid	10	50	90
Normal	2/5	$717,870	$1,717,710	$999,840	58%	$57.56	10%	4%	85%	-17%	150%	950%
	3/5		$1,273,500	$555,630	44%	$42.67	18%	6%	76%	-47%	86%	700%
	Weighted - 3/5		$1,300,980	$583,110	45%	$43.59	18%	5%	77%	-43%	93%	700%
	CAP - 5 days - no forward		$1,164,480	$446,610	38%	$39.02	21%	8%	71%	-50%	63%	467%
	Regression		$1,630,740	$912,870	56%	$54.64	9%	6%	85%	-11%	132%	600%
Threshold (Relative)	2/5	$403,260	$1,670,160	$1,266,900	76%	$55.96	4%	4%	93%	31%	325%	1380%
	3/5		$1,229,040	$825,780	67%	$41.18	9%	5%	86%	-20%	211%	1000%
	Weighted - 3/5		$1,254,600	$851,340	68%	$42.04	9%	5%	86%	-14%	221%	1000%
	CAP - 5 days - no forward		$1,119,480	$716,220	64%	$37.51	10%	7%	83%	-22%	167%	810%
	Regression		$1,588,380	$1,185,120	75%	$53.22	3%	5%	91%	44%	286%	1033%
Threshold (Absolute)	2/5	$343,770	$1,420,740	$1,076,970	76%	$47.61	5%	19%	76%	-24%	181%	1300%
	3/5		$993,840	$650,070	65%	$33.30	8%	26%	66%	-62%	92%	889%
	Weighted - 3/5		$1,019,760	$675,990	66%	$34.17	9%	25%	66%	-54%	100%	900%
	CAP - 5 days - no forward		$918,780	$575,010	63%	$30.79	8%	33%	60%	-52%	112%	850%
	Regression		$1,371,870	$1,028,100	75%	$45.97	2%	27%	71%	14%	233%	1313%
Graduated (Relative)	2/5	$522,228	$1,706,737	$1,184,509	69%	$57.19	7%	0%	93%	17%	234%	831%
	3/5		$1,251,031	$728,803	58%	$41.92	15%	0%	85%	-25%	146%	596%
	CAP - 5 days - no forward		$1,152,937	$630,709	55%	$38.63	20%	0%	80%	-32%	107%	489%
	Regression		$1,624,927	$1,102,699	68%	$54.45	8%	0%	91%	7%	182%	613%
	Weighted - 3/5		$1,277,699	$755,471	59%	$42.81	15%	0%	85%	-19%	148%	598%
Graduated (Absolute)	2/5	$601,839	$1,771,530	$1,169,691	66%	$59.36	8%	0%	92%	9%	188%	692%
	3/5		$1,274,659	$672,820	53%	$42.71	17%	0%	82%	-30%	110%	482%
	CAP - 5 days - no forward		$1,181,466	$579,626	49%	$39.59	24%	0%	76%	-38%	70%	381%
	Regression		$1,704,529	$1,102,690	65%	$57.12	12%	0%	88%	-4%	135%	486%
	Weighted - 3/5		$1,308,029	$706,190	54%	$43.83	17%	0%	83%	-24%	113%	489%
Monthly	2/5	$774,600	$1,463,280	$688,680	47%	$49.03	20%	4%	76%	-50%	114%	500%
	3/5		$1,014,360	$239,760	24%	$33.99	31%	6%	63%	-79%	50%	350%
	CAP - 5 days - no forward		$920,640	$146,040	16%	$30.85	40%	8%	52%	-78%	13%	267%
	Regression		$1,365,720	$591,120	43%	$45.76	23%	7%	71%	-50%	67%	317%
	Weighted - 3/5		$1,043,700	$269,100	26%	$34.97	30%	5%	64%	-76%	50%	350%

Table C-3: Payment Accuracy Tables for Unadjusted Baselines during the 11AM – 6PM Event Window Scenario 3: Targeted Enrollment (20% Impacts, 120,000 Customers, 5 Events)

Payment Method	Baseline	Total Payment				$/kW-year	Payment Accuracy			Individual Customer Payment Error Pctiles		
		$ Actual	$ Estimated	$ Error	% of Payment in Error		Under-paid	Paid Correctly	Over-paid	10	50	90
Normal	2/5	$768,300	$1,095,765	$327,465	30%	$36.72	22%	4%	74%	-33%	56%	317%
	3/5		$841,620	$73,320	9%	$28.20	34%	6%	60%	-55%	22%	227%
	Weighted - 3/5		$858,795	$90,495	11%	$28.78	33%	6%	61%	-53%	24%	220%
	CAP - 5 days - no forward		$771,375	$3,075	0%	$25.85	44%	7%	49%	-59%	0%	163%
	Regression		$1,043,805	$275,505	26%	$34.98	27%	6%	67%	-34%	35%	217%
Threshold (Relative)	2/5	$768,300	$1,081,530	$313,230	29%	$36.24	22%	4%	74%	-36%	55%	317%
	3/5		$826,455	$58,155	7%	$27.69	35%	6%	59%	-57%	21%	225%
	Weighted - 3/5		$844,755	$76,455	9%	$28.31	33%	6%	61%	-54%	23%	220%
	CAP - 5 days - no forward		$756,750	-$11,550	0%	$25.36	44%	7%	49%	-60%	0%	163%
	Regression		$1,025,340	$257,040	25%	$34.36	28%	6%	66%	-36%	33%	214%
Threshold (Absolute)	2/5	$548,385	$933,480	$385,095	41%	$31.28	16%	15%	69%	-47%	62%	667%
	3/5		$676,740	$128,355	19%	$22.68	24%	20%	56%	-70%	15%	417%
	Weighted - 3/5		$695,445	$147,060	21%	$23.30	23%	20%	58%	-67%	22%	425%
	CAP - 5 days - no forward		$622,365	$73,980	12%	$20.85	25%	27%	48%	-74%	6%	350%
	Regression		$891,420	$343,035	38%	$29.87	13%	25%	62%	-36%	59%	460%
Graduated (Relative)	2/5	$522,284	$1,095,278	$572,994	52%	$36.70	12%	0%	88%	-7%	124%	433%
	3/5		$827,030	$304,746	37%	$27.71	21%	0%	79%	-35%	71%	310%
	CAP - 5 days - no forward		$760,258	$237,974	31%	$25.47	29%	0%	70%	-40%	43%	237%
	Regression		$1,044,958	$522,674	50%	$35.01	13%	0%	87%	-10%	86%	304%
	Weighted - 3/5		$844,104	$321,820	38%	$28.28	20%	0%	79%	-33%	73%	308%
Graduated (Absolute)	2/5	$718,714	$1,143,665	$424,951	37%	$38.32	19%	0%	81%	-26%	77%	325%
	3/5		$849,448	$130,734	15%	$28.46	32%	0%	68%	-52%	35%	224%
	CAP - 5 days - no forward		$787,986	$69,272	9%	$26.40	43%	0%	56%	-54%	11%	159%
	Regression		$1,099,550	$380,836	35%	$36.84	26%	0%	74%	-31%	41%	210%
	Weighted - 3/5		$871,538	$152,824	18%	$29.20	31%	0%	69%	-48%	37%	226%
Monthly	2/5	$782,355	$994,320	$211,965	21%	$33.32	27%	4%	68%	-50%	46%	223%
	3/5		$728,070	-$54,285	0%	$24.40	42%	6%	53%	-73%	10%	150%
	CAP - 5 days - no forward		$668,400	-$113,955	0%	$22.40	54%	6%	40%	-75%	-13%	100%
	Regression		$941,985	$159,630	17%	$31.56	37%	6%	57%	-55%	17%	133%
	Weighted - 3/5		$748,425	-$33,930	0%	$25.08	41%	5%	54%	-71%	11%	150%

Appendix D Overview of PTR Programs and Pilots

This appendix contains a brief summary of the purpose, the various rate and technology options examined, recruitment methods (e.g., opt-in, opt-out, etc.), timing and other high level characteristics of PTR pilots and programs that have been implemented in other jurisdictions.

D.1 Pilot and Program Overview

Table D-1 provides a high level summary of key characteristics of the pilots included in this appendix. To our knowledge, the list shown in the table includes all of the pilots that have been conducted in North America that included PTR either as the only rate tested or as one tested in conjunction with other rate options. Evaluations of these pilots have produced estimates of the demand response that resulted from the specific tariffs tested. In addition, several recent operational or "pre-launch" pilots have been implemented to test operational readiness in anticipation of a larger roll out of PTR.

Previous pilots vary with respect to location, date, duration, prices tested, sample sizes, experimental design and other factors. They also vary with respect to enrollment strategies (e.g., opt-in versus opt-out). The first six pilots listed in Table D-1 recruited customers on an opt-in basis while the last two defaulted customers onto PTR and gave them the opportunity to opt out. In some cases, the pilots include enabling technology such as Energy Orbs and AC cycling devices for some test cells. The magnitude of the PTR credits varied as well, with rebates ranging from as low as $0.35/kWh to as high as $1.75/kWh.

Table D-1: Overview of PTR Pilots

Enrollment	Utility	State	Tested vs. CPP?	Year	Rebate/Adder ($/kWh)	PTR Sample sizes
Opt-in	Anaheim PU	CA	No	2005	$0.35	126
	BG&E	MD	Yes	2008 - 2012	Low: $1.16 High: $1.75	Low: n=126 High: n=127 Tech: n=509
	CL&P	CT	Yes	2009	Low: $1.16 High: $1.75	Low: n=108 High: n=100 Tech: n=174
	Consumer's Energy	MI	Yes	2010	Standard; $0.50	152
	PEPCO (PowerCents DC*)	DC	Yes**	2008-2009	Standard: $0.66 Low income: $0.83	262
	Ontario Energy Board (Ottawa Hydro)	ON	Yes	2007	$0.30	125
Default	Commonwealth Edison (ComED)	IL	Yes	2010	$1.74	Standard: 225 Tech: 1075
	San Diego Gas & Electric (SDG&E)	*CA*	*No*	*2011*	*Standard: $0.75* *Tech: $1.25*	*Standard: 2,900* *Tech: 100*

*All participants were offered smart thermostats; 1/3 of them took up the offer and had their units automatically controlled during events.

**There were no low income CPP customers for the final analysis. PTR v. CPP comparison is only valid for non-low income customers.

D.1.1 Anaheim Public Utilities[19]

APU is a municipal utility in Southern California currently serving more than 345,000 residential customers. The pilot was conducted during summer 2005 and was marketed under the name, Spare the Power Days Rebate Program. Control and treatment groups were selected at random and the treatment group was recruited into the study using a combination of direct mail and telephone calls.[20] Customers were not told that they could opt out of the program and less than five of those contacted by phone chose not to participate. Sample sizes were small, with only 52 customers in the control group and 71 in the treatment group. The PTR credit was $0.35/kWh and the event hours were from noon to 6 PM. Events were called one day ahead. Twelve events were called during the summer, which ran from June through mid-October.

[19] Wolak, 2006.

[20] Note that this is not a randomized control trial design that controls for selection bias in the enrollment process. Controlling for selection bias would require making random assignment to treatment and control after customers decided to join the program (a recruit and deny or recruit and delay design).

D.1.2 Baltimore Gas & Electric[21]

BGE is a utility serving more than one million customers in Baltimore and central Maryland. The first year of the Smart Energy Pricing Pilot was conducted during summer 2008. BGE's pilot tested two PTR credit levels; PTR Low and PTR High. It also tested a CPP rate. Treatment customers were selected from two previously created sample groups. These customers were recruited into the study through direct mail and telephone calls. Recruitment information only offered a single rate to each customer. The control group was comprised of customers in the sample groups who were not able to be contacted or recruited. This almost certainly introduces selection effects into the estimation process that could bias the impact estimates. Ultimately, there were approximately 260 customers in the CPP treatment group, 380 in the PTRL group, 380 in the PTRH group, and 350 in the control group.

The rebates for the PTR groups were $1.16/kWh for PTRL and $1.75/kWh for PTRH. CPP customers paid $1.16/kWh on top of normal prices during critical peak hours. Additionally, BGE tested the impacts of an Energy Orb and a switch for cycling air conditioners alongside the dynamic pricing options. Event hours were 2 PM to 7 PM and were called a day ahead. Twelve events were called during the initial pilot period, which ran from June through September.

After evaluating results from the first summer of the pilot, BGE continued the pilot with only PTR customers. The pilot has been in place for a total of four years now, making it the longest running pilot in the industry.

D.1.3 Connecticut Light & Power

CL&P is an electric utility located in New England currently serving approximately 1.2 million customers. The pilot was named the Plan-It Wise Energy Pilot and was conducted in summer 2009. CPP was tested in addition to PTR. Control and treatment groups were randomly selected, and treatment groups were recruited into the study using direct mail and telephone calls. There were approximately 380 customers in the PTR treatment group, 370 in the CPP treatment group and 140 in the control group. PTR and CPP rebates and rates were split into Low and High groups. PTR credits were $0.78/kWh for PTRL and $1.74/kWh for PTRH. Customers on the CPP rate paid $0.86/kWh for CPPL and $1.80/kWh for CPPH during critical peak hours. CL&P also tested the impacts of four enabling technologies which included In-Home Displays, Energy Orbs, Smart Thermostats and AC Control Switches. CL&P's event hours were 2 PM to 6 PM and events were called a day ahead. Ten event days were called during the pilot, which ran from June through September.

D.1.4 Consumers Energy[22]

CE is a public utility located in Michigan currently serving nearly 6.5 million residential customers. The program, named the Personal Power Plan Pilot, was conducted in Jackson, Michigan during summer 2010. PTR was tested alongside CPP, and enabling technology was tested with CPP only. Control and treatment groups were randomly selected, and the treatment groups were recruited using direct mail and follow up telephone calls. Recruitment information only described one rate structure. There were approximately 150 customers in the PTR treatment group, 220 customers in the CPP treatment group and 320 customers in the control group. The PTR credit was $0.50/kWh and the CPP critical peak

[21] Faruqui and Sergici, 2009.

[22] Faruqui, Sergici and Lamine, 2012.

price was $0.69/kWh. Events were called a day ahead and ran from 2 PM to 6 PM. Six event days were called during the pilot, which ran from July through September.

D.1.5 Ontario Energy Board[23]

The Ontario Energy Board conducted the Smart Price Pilot during summer 2006 through winter 2007. The pilot was marketed under the name PowerWise and was implemented in the Hydro Ottawa service territory, a municipal electricity distribution company currently serving approximately 300,000 customers. PTR was tested alongside a CPP rate. Control and treatment groups were randomly selected from Hydro Ottawa's territory. The PTR and CPP treatment groups were recruited using direct mail and customers were only included in the pilot if they returned a confirmation form included in the mailing. With a 25.5% response rate, there were ultimately 125 customers each in the 2 treatment groups and the control group. The PTR credit was $0.30/kWh and the CPP all-in rate was $0.30/kWh during event hours. To facilitate billing, PTR customers did not receive rebates until the end of the pilot. In this pilot, events were called a day ahead and ran for three to four hours sometime between 11 AM to 5 PM depending on the day. Seven events were called during the pilot period, which ran from June 2006 through February 2007.

D.1.6 Pepco[24]

Pepco is an investor-owned utility currently serving nearly 800,000 customers in Washington, D.C. and parts of Maryland. From summer 2008 through summer 2009, a commission-appointed collaborative conducted the PowerCentsDC pilot in the District of Columbia. The pilot tested PTR along with CPP, and included enabling technology in the form of a smart thermostat. Treatment and control customers were selected at random but were not randomly assigned to treatment and control groups after selection as would have been done in a randomized control trial (RCT) pilot. Treatment customers were recruited through direct mail and were only included if they returned the confirmation form included in the mailing. With a response rate of 7.4%, the PTR treatment group had approximately 320 customers. The CPP treatment group had a lower response rate of 6.5% and 230 customers. The control group had 378 customers. The PTR credit was $0.66/kWh and the CPP all-in price during event hours was $0.78/kWh. Events were called a day ahead and event hours were from 2 PM to 6 PM. There were 6 events during summer 2008, 3 events in January 2009, and 12 events in summer 2009. The pilot ran from July 2008 through October 2009.

D.1.7 Commonwealth Edison[25]

ComEd is an electric utility located in Northern Illinois currently serving approximately 3.8 million customers. The Customer Application Program was conducted in 2010 and 2011 and tested CPP in addition to PTR. The control and treatment groups were selected at random, and customers were automatically enrolled into one specific pricing structure. This was the first default pilot design that was implemented in the industry and one of only two that has been conducted. It is a good example of an RCT design. There were 450 customers in the control group, 980 in the PTR treatment group and 1,900 in the CPP treatment group. The PTR rebate was $1.74/kWh while the CPP customers paid an extra $1.74/kWh during event hours on top of their normal rates. Both the PTR credits and CPP

[23] Wolak, 2007.

[24] Wolak, 2010.

[25] EPRI 1023644, October 2011.

prices were layered on top of an underlying hourly energy price. Events were called a day ahead and event hours were from 1 PM to 5 PM. There were seven events called during the pilot period, which ran from June 2010 to May 2011.

D.1.8 Pre-launch Pilots

In addition to the pilots summarized above, which were implemented primarily to test the load impact associated with PTR and other rate options, several utilities that are planning to implement PTR on a large scale have recently conducted small, pre-launch pilots designed primarily to test operational readiness. In 2012, Delmarva Power in Delaware and Pepco in Maryland implemented pre-launch pilots with 6,900 and 5,000 customers, respectively. These pilots did not involve deployment of treatment and control groups. However, it is still possible to estimate impacts because of the "on-off" nature of the incentive. As of this writing, no publicly available estimates of load impacts were available.

D.2 Summary of Load Impacts

Figure D-1 shows the estimated load impacts for PTR from each of the pilots for which results are available, as well as the impacts for CPP where valid comparisons can be made. The figure also shows the PTR incentives (or equivalent CPP prices) that underlie the impacts in each pilot. The values in Figure D-1 are only for treatment cells that do not include enabling technology.

As seen, the impacts vary significantly across pilots. For PTR, the lowest impact was found for the low incentive treatment ($0.78/kWh) in the CL&P pilot and the highest impact was for the high incentive treatment ($1.75/kWh) in the BGE pilot. There is no strong correlation between prices and impacts across pilots, but this is not surprising given the significant differences in population characteristics, average electricity prices and other factors that vary across studies. For example, it is not appropriate to compare impacts for the Anaheim pilot, which was conducted in a very mild, Southern California climate zone with a relatively low saturation of air conditioning and low humidity, with impacts for BGE, where air conditioning saturations, temperature and humidity are all much higher. Similarly, impacts for pilot participants in Ontario may differ from other utilities because more than 80% of program participants have college or graduate degrees, which is much greater than in any other pilot. For these reasons, comparing impacts across rate treatments within pilots is much more valid and insightful. For studies that have multiple incentive levels, higher incentives produce higher impacts, although the differences are not terribly large in the two pilots where multiple incentive levels were tested.

It should be noted that the impacts in Figure D-1 for BGE are from summer 2008, the first year of their pilot. BGE dropped the CPP rate after 2008 but continued to offer PTR to participants in subsequent years. Impacts have persisted over time and in fact have increased slightly (although it is unclear from available documentation whether any changes over time are statistically significant).[26]

[26] See Harbaugh, 2011.

Figure D-1: Load Impacts by Rate Type and Pilot for Treatments with No Enabling Technology

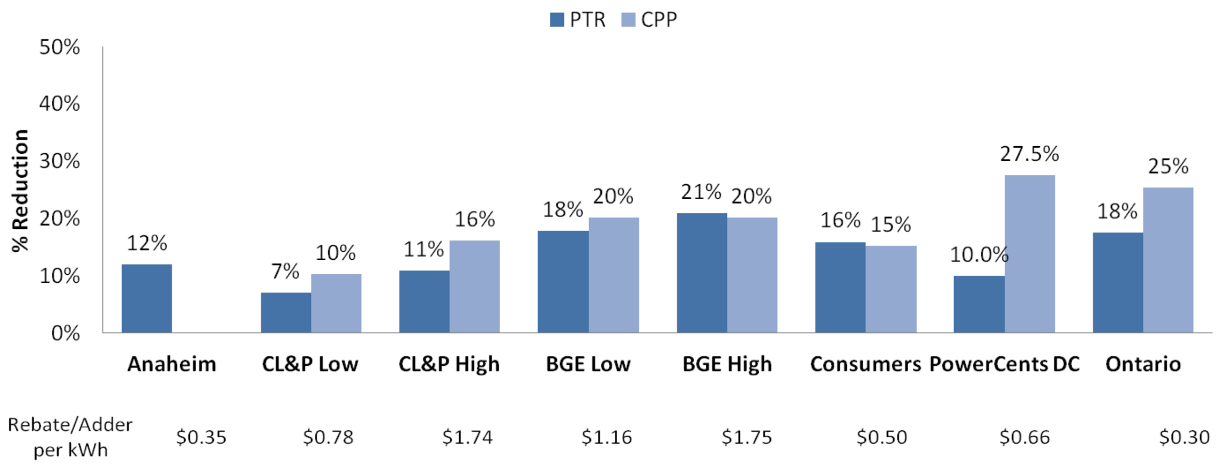


Figure D-2 compares the load impacts for PTR and CPP for treatments in which enabling technology is deployed. In most cases, PTR impacts are less than CPP impacts. Overall, the difference between the two rate options with enabling technology is roughly 25%.

Figure D-2: Load Impacts by Rate Type and Pilot for Treatments with Enabling Technology

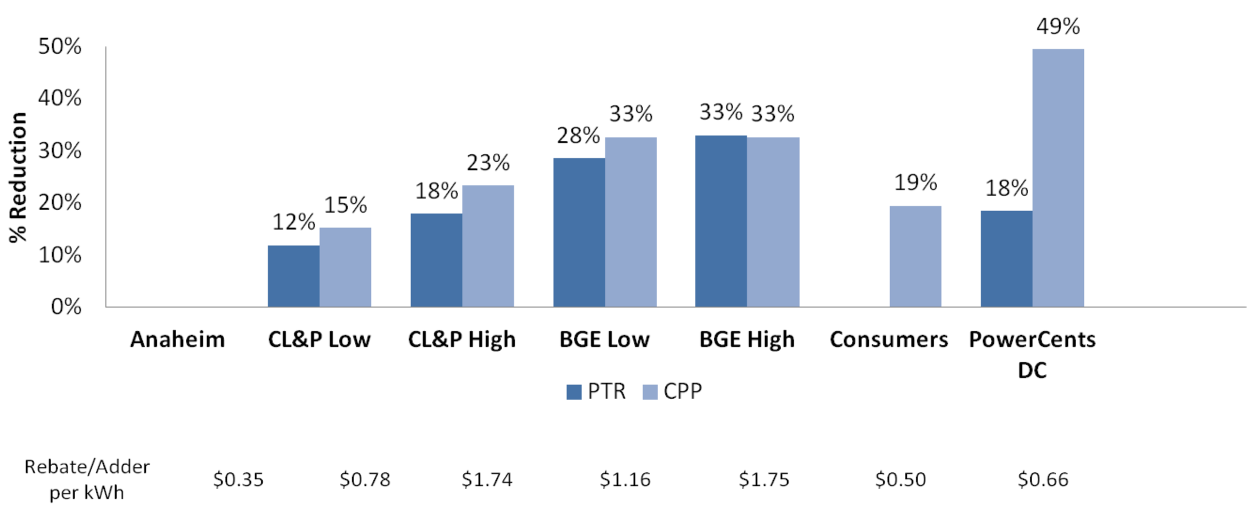
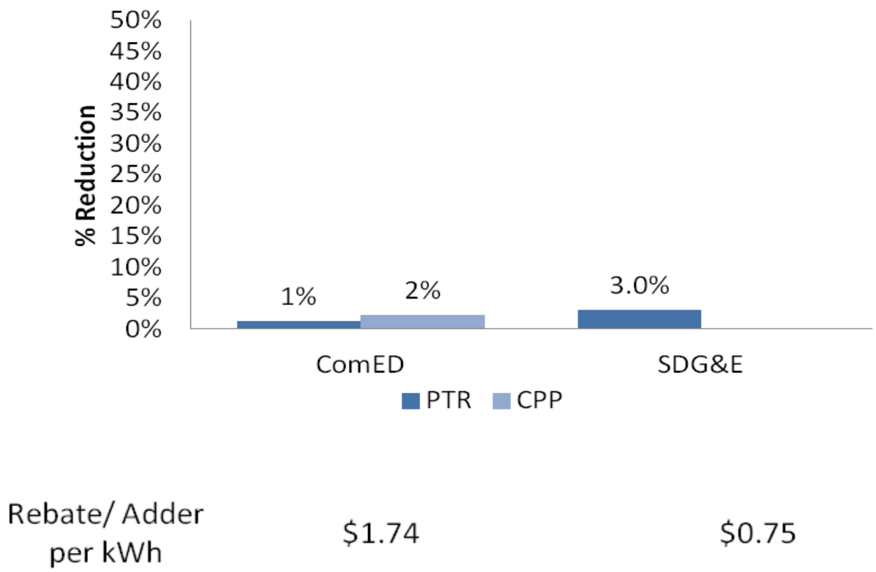


Figure D-3 shows the average load impact for the two default pilots. The average impact for all customers exposed to the PTR incentive was quite low in both studies. Unlike with several of the opt-in pilots, where control group selection is less than ideal and may not adequately control for selection effects,[27] these default pilots are well designed, randomized control trials in which treatment and control customers are very well matched. We have no reason to believe that these average impacts do not accurately reflect what might be achieved for a broader population within these service territories for rates similar to those tested.

[27] See George, July 2010, Appendix A for a discussion of concerns about control group selection and other issues associated with four of the pilots: BGE, CL&P, Ontario and PowerCentsDC.

Figure D-3: Load Impacts for Default PTR Pilots



It should be noted that Oklahoma Gas & Electric (OG&E) has also implemented a pilot, although one that did not test PTR. OG&E's Smart Study Together pilot tested a variable peak tariff (VPP) and a CPP/TOU tariff during the summers of 2010 and 2011. The VPP tariff had four different day types across which peak period prices varied as follows:

- Low peak and off-peak price of $0.045/kWh;
- Standard day peak price of $0.11/kWh;
- High day peak price of $0.23/kWh; and
- Critical day and event price of $0.46/kWh.

In 2011, OG&E called seven event days. The pilot also included treatment cells that variously gave customers access to a web portal, an in-home display (IHD) and/or installation of a PCT that eligible customers (e.g., those with central air conditioning) could set to automate response during the peak period on high priced days. Impacts varied across treatment groups. Those with access only to an IHD or the web portal gave an average reduction of roughly 6% across the event days in 2011 while those with a PCT produced an average reduction of roughly 27%. Similar impacts were also obtained for the CPP/TOU treatment customers.

D.3 Bibliography

The following reports and presentation discuss PTR pilots and programs. Some are cited above.

EPRI. *The Effect on Electricity Consumption of the Commonwealth Edison Customer Application Program Pilot: Phase 1*, Electric Power Research Institute, Palo Alto, CA. EPRI 1022703. April 2011.

EPRI. *The Effect on Electricity Consumption of the Commonwealth Edison Customer Applications Program: Phase 2 Final Analysis*, Christensen, Electric Power Research Institute, Palo Alto, CA. EPRI 1023644. October 2011.

Faruqui, Ahmad, Sergici, Sanem. *Household Response to Dynamic Pricing of Electricity: A Survey of Seventeen Pricing Experiments*, The Brattle Group, 2008.

Faruqui, Ahmad. *Demand Response: A Long View*, The Brattle Group, 2008.

Faruqui, Ahmad, Sergici, Sanem. *The Power of Experimentation: New Evidence on Residential Demand Response*, The Brattle Group, 2008.

Faruqui, Ahmad. *Impact Evaluation of CL&P's Plan-It Wise Energy Program*, The Brattle Group, 2009.

Faruqui, Ahmad, Sergici, Sanem. *BGE's Smart Energy Pricing Pilot Summer 2008 Impact Evaluation*, The Brattle Group, 2009.

Faruqui, Ahmad, Hledik, Ryan, Sergici, Sanem. *Rethinking Prices: The Changing Architecture of Demand Response in America*, Public Utilities Fortnightly, 2010.

Faruqui, Ahmad, Sergici, Sanem, Akaba, Lamine, *Dynamic Pricing in a Moderate Climate: The Evidence from Connecticut*, The Brattle Group, 2012.

Faruqui, Ahmad, Sergici, Sanem, Akaba, Lamine. *Dynamic Pricing of Electricity for Residential Customers: The Evidence from Michigan*, The Brattle Group, 2012.

George, Stephen, Faruqui, Ahmad. *Impact Evaluation of the California Statewide Pricing Pilot*, Charles River Associates, 2005.

George, Stephen, Bode, Josh, Perry, Mike, Mayer, Zach. *2009 Load Impact Evaluation for Pacific Gas and Electric Company's Residential SmartRate™—Peak Day Pricing and TOU Tariffs and SmartAC Program—Volume 1: Ex Post Load Impacts*, Freeman, Sullivan & Co., 2009.

George, Stephen, Bode, Josh, Perry, Mike, Goett, Andrew. *2009 Load Impact Evaluation for Pacific Gas and Electric Company's Residential SmartRate™—Peak Day Pricing and TOU Tariffs and SmartAC Program—Volume 2: Ex Ante Load Impacts*, Freeman, Sullivan & Co., 2009.

George, Stephen. *Overview of Price Driven Demand Response and Recommendations for Implementing a Pricing Pilot by Central Maine Power.* Filed by Central Maine Power in conjunction with Docket No. 2010-132: Investigation into Central Maine Power Company's AMI related Programs. Filed July 30, 2010.

Global Energy Partners. *OG&E Smart Study Together Impact Results. Final Report – Summer 2011.* February 29, 2012.

Harbaugh, Wayne. *Smart Grid/Dynamic Pricing Update for BGE.* Presented at the EEI Rates and Regulatory Affairs Committee meeting. September 13, 2011.

Hindes, Cheryl. *Introducing the Smart Grid*, Presentation of Baltimore Gas and Electric Company, 2009.

Hindes, Cheryl. *Smart Energy Pricing: "Customers are making it work,"* Presentation of Baltimore Gas and Electric Company, 2010.

IEE Whitepaper. *The Impact of Dynamic Pricing on Low Income Customers*, 2010.

King, Chris. *PowerCentsDC Program Final Report*, EMeter Strategic Consulting, 2010.

Morgan, Rick. *Smart Pricing for the Smart Grid*, Federal Briefing Slides, 2007.

Strapp, James, King, Chris, Talbott, Sharon. *Ontario Energy Board Smart Price Pilot Final Report*, IBM and EMeter, 2007.

Strategic Consulting. *Ontario Energy Board Smart Price Pilot Final Report*, 2007.

Strategic Consulting. *PowerCentsDC™ Program Interim Report*, 2009.

Summit Blue Consulting, LLC. *Final Report for the MyPower Pricing Segments Evaluation*, 2007.

Wolak, Frank A. *An Experimental Comparison of Critical Peak and Hourly Pricing: The PowerCentsDC Program*, 2010.

Wolak, Frank A. *Load Impact and Conservation Effect Analytical Model: Ontario Smart Price Pilot,* 2007.

Wolak, Frank A. *Residential Customer Response to Real-Time Pricing: The Anaheim Critical-Peak Pricing Experiment*, 2006.