LBNL-61090

Ernest Orlando Lawrence
Berkeley National Laboratory

Residential Demand Response Evaluation: A Scoping Study

C.K. Woo, K. Herter

June 2006

LBNL-61090
Collaborative Report

Residential Demand Response Evaluation: A Scoping Study

C. K. Woo
Energy and Environmental Economics, Inc. (E3)
353 Sacramento Street, Suite 1700, San Francisco CA 94111

Karen Herter
Lawrence Berkeley National Laboratory
1 Cyclotron Road, MS90-3111, Berkeley, CA 94720

June 2006

This work described in this report was coordinated by the Demand Response Research Center and funded by the California Energy Commission, Public Interest Energy Research Program, under Work for Others Contract No. 500-03-026 and by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

PREFACE

The Public Interest Energy Research (PIER) Program supports public interest energy research and development that will help improve the quality of life in California by bringing environmentally safe, affordable, and reliable energy services and products to the marketplace.

The PIER Program, managed by the California Energy Commission (Commission), annually awards up to $62 million to conduct the most promising public interest energy research by partnering with Research, Development, and Demonstration (RD&D) organizations, including individuals, businesses, utilities, and public or private research institutions.

PIER funding efforts are focused on the following six RD&D program areas:

- Buildings End-Use Energy Efficiency
- Industrial/Agricultural/Water End-Use Energy Efficiency
- Renewable Energy
- Environmentally-Preferred Advanced Generation
- Energy-Related Environmental Research
- Energy Systems Integration

What follows is the final report for the Residential DR Scoping Study Project, 500-03-026 Task 4J, conducted by Energy and Environmental Economics (E3). The report is entitled “Residential Demand Response Evaluation Scoping Study.” This project contributes to the Energy Systems Integration Program.

For more information on the PIER Program, please visit the Commission's Web site at: http://www.energy.ca.gov/research/index.html or contact the Commission's Publications Unit at 916-654-5200.

ACKNOWLEDGEMENTS

This work described in this report was coordinated by the Demand Response Research Center and funded by the California Energy Commission's Public Interest Energy Research (PIER) Program, under Work for Others Contract No. 500-03-026 and by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

The authors wish to thank the members of our Technical Advisory Group: David Hungerford and Adrienne Kandel of the California Energy Commission, Ed Vine of Lawrence Berkeley National Lab, and Susan McNichols of Pacific Gas and Electric Company for their input, suggestions and comments.

ABSTRACT

The primary goals of this scoping study were to (1) summarize existing methods for estimating demand response, (2) evaluate these methods' abilities to accurately estimate residential demand response for the purpose of program evaluation, (3) recommend a preferred approach, and (4) outline any remaining knowledge gaps. This study was motivated by the CPUC directive (D.05-11-009) of developing measurement and evaluation protocols for demand response.

Our evaluation considers both day-matching and regression techniques, outlining the following alternative methods: (1) prior-day averaging, (2) weather-matching techniques, (3) regression-based load profile comparison, and (4) econometric demand analysis. Based on a review of these methods for evaluating demand response, we find that customer-specific regression analysis is likely to give accurate, transparent and intuitive results. Depending on program requirements, this method can be modified to estimate hourly demand response before, during and after events, providing hourly kW response results and load profiles.

Beyond basic demand response estimation, several issues need to be addressed before a practical method for residential demand response program evaluation can be determined. Among them are the ability to evaluate multiple events on consecutive days, an understanding of how advance notification affects demand response, and incorporation of considerations affecting the extrapolation of results from a voluntary pilot to a large-scale program.

1. INTRODUCTION

In April 2005, the Partner Planning Committee of the Demand Response Research Center (DRRC) agreed on establishing the “Value of Demand Response” as their top research priority. To find the value of residential demand response, however, it is necessary to accurately measure customer load impacts resulting from price or reliability triggered demand reduction events (Herter, McAuliffe and Rosenfeld, 2006; Horowitz and Woo, 2006). Hence, the primary focus of this scoping study is to identify methods that, when applied to actual customer load data,[1] can produce unbiased load impact estimates.[2]

After fully incorporating the valuable input from the project's Technical Advisory Group (TAG),[3] this final report summarizes existing methods for estimating demand response, evaluates these methods' abilities to accurately estimate residential demand response for

[1] The recorded load data may come from a demand response experiment or an on-going program.
[2] Thus, the study is not a review of engineering estimates for a kW change due to a higher thermostat setting on a hot summer day. Nor is the study an assessment of the ability of a particular method to accurately predict demand response to program designs as yet to be implemented. Such an assessment is beyond the scope of this study.
[3] The TAG team includes: Adrienne Kandel and David Hungerford of the California Energy Commission, Susan McNicoll of Pacific Gas and Electric Company, and Ed Vine of Lawrence Berkeley National Laboratory. The team's detailed comments have greatly improved the report's content and exposition.

the purpose of program evaluation (rather than individual customer response), recommend a preferred approach, and outlines any remaining knowledge gaps. These results help meet the CPUC directive (D.05-11-009) of developing measurement and evaluation protocols for demand response. They also contribute to the DRRC's effort in refining the cost-effectiveness tests in the Standard Practice Manual (Baskette, Horii, Kollman and Price, 2006).

We define demand response as the difference between (a) the unperturbed load profile absent the event; and (b) the perturbed load profile as a result of the event. We refer to the unperturbed load profile in (a) as the "baseline" load profile. Since the perturbed profile in (b) is actually observed, accurate demand response estimation predicates on accurate estimation of the baseline loads.

The common methods of demand response estimation include:

- **Prior-days averaging** techniques, which compare the load profile on an event day to the average profile over multiple days prior to the event. The average non-event profile is at times adjusted to match the actual loads in the morning before the event hours.

- **Weather-matching** techniques, which compare the load profile on one or more event days to the average profile over non-event days with similar weather characteristics.

- **Regression-based load comparison,** which entails estimating customer-specific hourly load regressions that incorporate the effects of time, weather and event variables.

- **Econometric demand models** based on a system of electricity demand equations, typically derived from utility-maximization behavior of electricity consuming households.[4]

A detailed summary and critique of these methods, along with an assessment of critical knowledge gaps, will be presented in this summary report. The mathematical details of each method are provided in an appendix.

Based on the findings detailed below, we conclude that a regression-based load comparison approach is likely to provide a reasonable estimate of residential demand response to a utility-triggered event. However, critical knowledge gaps in the effort to evaluate residential demand response remain. By no means exhaustive, the list of gaps includes: (a) validation of the accuracy of response estimates; (b) modeling the effect of consecutively triggered events on response estimates; (c) modeling the effect of a day-ahead notice in the presence of public appeal for voluntary load curtailment; and (d) estimating aggregate load impact under a voluntary tariff. Hence, we urge the DRRC to continue its support of demand response estimation as part of its on-going research.

[4] The popular double-log and linear demand equations are consistent with utility maximization (Hausman, 1981). As noted by a TAG member, however, a demand model may be empirically based, purely driven by price/quantity data.

2. CHALLENGES

A sound method should overcome challenges frequently encountered in any demand response estimation: ease of implementation, accuracy, and transparency. These challenges are discussed below.

2.1 Ease of implementation

When choosing a method for evaluating individual customer demand response, simple calculation is a top priority, since both customers and program administrators are expected to make use of the method (Goldberg and Agnew, 2003). In the case of program evaluation, however, ease of implementation is less critical, as it is conducted by those familiar with analysis techniques. At the same time, a complex method requiring substantial time and money for expert analysis is undesirable. Ideally, the method will be as simple to use as possible without sacrificing accuracy or usefulness. The following questions address some of the challenges a good method should overcome.

Can the method easily handle voluminous hourly datasets?

Demand response estimation typically uses hourly load and climate data over multiple months or years from multiple customers dispersed across geographical zones. These datasets can quickly become rather cumbersome. For example, the recent 15-month pricing experiment in California generated about 22 million hourly load data values and over 600,000 temperature values for the roughly 2,000 participating customers in four climate zones.

Such large datasets present a challenge to the identification and quantification of customer response. Hence, a good approach must be computationally efficient, overcoming the "too much data" problem. It should do so without imposing potentially restrictive assumptions on the data, such as might be done by averaging over like groups (e.g. all customers within a climate zone).[5]

Can the method easily estimate load shifting and spill-over effects?

Part or all of residential response to an event can be the result of shifting loads from the event hours to (a) neighboring hours in the event day, (b) hours in the preceding day after receiving the advance notice; or (c) hours in the following day. Some household activities, like clothes and dish washing, can be done earlier or later than planned with little or no inconvenience. Even air-conditioning can be activated in the hours before an event to pre-cool the house, allowing reduced air-conditioning load during event hours. Where air-conditioning is forgone during events, increased indoor temperatures are likely to raise air-conditioner utilization in the post-event hours.

[5] A TAG member notes that industries like banking and telecommunication daily handle large data files, with little or no problem in real-time processing bills or extracting data. However, data analysis is not the same as data processing and retrieval. Personal communication with a vice president of the E-commerce division of a very large bank confirms that it is difficult to accurately analyze the pattern of customer credit card purchases using very large data files.

Prediction of load shifts to hours outside the event period is important for enabling system operators to simultaneously shift supply to those hours. It is also important for estimating the effects of tariffs or programs on emissions, because of the time-varying difference in emissions rates of generation sources. Hence, a thorough understanding of the effects of price or reliability events requires an approach that can calculate these spill-over effects, which have not been fully investigated in the extant literature reviewed below.

2.2 Accuracy

The second set of challenges arises from the nature of residential demand response. A sound method should account for the possibility that demand response may depend on such factors as weather, timing, and customer demographics; otherwise, the resulting estimates may be biased, a consequence of under-specification. When modeling these dependencies, however, the approach must not impose excessive computational time and complexity that can render it impractical for frequent applications.[6]

Does the method control for the effects of time on response?

Customer responses are time dependent. Residential loads vary substantially by time-of-day and day-of-week. They are also likely to have a seasonal pattern, or month-of-year effects. A good approach must be sufficiently rich to capture the kW variance due to the effects of time-of-day, day-of-week and month-of-year. Constraining responses to be time insensitive likely introduces estimation bias.

Does the method control for the effects of weather on response?

Residential response is also weather-dependent (Herter, McAuliffe and Rosenfeld, 2006). Any accurate model comparing baseline and perturbed loads must control for hourly weather effects. In addition, they should recognize effects of persistent extreme weather on household load: thermal loads can increase on subsequent days of a prolonged heat wave or cold spell due to building heat retention or loss. Thus, a customer's response might also increase or decrease as the extreme weather persists.[7]

Does the method account for heterogeneous response?

Customer responses are heterogeneous, reflecting diverse customer characteristics like income, house size, and geographic location. One may estimate a single model of

[6] A TAG member comments that the ultimate choice of a specific method is more driven by the application at hand (e.g., circuit load estimation vs. individual load estimation) than its computational time or complexity. Concurring with this comment, the discussion below shows that a load-regression approach is not the easiest computation among the alternatives considered. However, it is a reasonable method to accurately quantify the kW impact using actual customer load data, the primary focus of the scoping study.
[7] A TAG member notes that the core TOU rate structure of the CPP tariff may cause a permanent change in air-conditioning use, as indicated by demand studies of time of use pricing (Aigner, 1986). The load-regression approach described in the Appendix below can analyze if a permanent change has occurred, given suitable data that allows a comparison of (a) the kW profile under non-TOU rates, (b) the kW profile under TOU rates only; and (c) the kW profile under CPP rates with TOU core.

customer response with parameters that link response differences to variations in customer characteristics – for example, an income coefficient for the income variable, a house size coefficient for the house size variable, etc. The resulting model, however, implicitly asserts that heterogeneous responses can be sufficiently represented by these parameters that are not allowed to vary across customers.[8] A more accurate and flexible model would estimate the parameters for each customer.

Moreover, given the weather-dependency of demand response, a single model of customer response, despite the inclusion of parameters that control for weather effects, would overlook the fact that weather-dependent response estimates can in turn depend on individual customer characteristics. For example, a customer with central air-conditioning is likely to respond differently than one without central air-conditioning. While the single model can use cross-product terms to account for the interaction between weather and demographics, it can become so large that its parameters are hard to track and interpret.

2.3 Usefulness

The last challenges relate to the substantive question: are the results useful in program evaluation, design and acceptance?

Is the method transparent?

A good method should be transparent, producing estimates that are easy to understand and verify. A demand response estimate from a black box, even if valid, is less credible than one found under a direct and intuitive approach. A program design based on non-transparent results invite skepticisms, hampering its acceptance by stakeholders (e.g., regulatory staff, utility staff, consumer advocates, and ratepayers).

Are the results intuitive?

A good method should produce a direct measurement of a customer's demand response, strictly adhering to its definition of kW difference between the customer's baseline and the perturbed load profiles. If the baseline profile comes from a complicated simulation exercise,[9] the load impact results are less convincing than those found from a simple one.

3. POTENTIAL METHODS FOR EVALUATING AGGREGATE RESIDENTIAL DEMAND RESPONSE

Our evaluation considers both day-matching and regression techniques, outlining the following alternative methods: (1) prior-day averaging, (2) weather-matching techniques, (3) regression-based load profile comparison, and (4) econometric demand analysis. The

[8] One may estimate a random coefficient model to reflect taste variance. But such an approach is computationally difficult and non-transparent.
[9] An example is the simulation of baseline profiles via price scenarios created as input data to a large system of estimated demand equations.

resulting discussion applies to the evaluation of hourly data produced by either the price events of a dynamic rate or the reliability events of a dispatchable load control program.

3.1 Day-matching techniques

These techniques assume the existence of one or more "non-event" days whose hourly load values can be averaged to provide the baseline load profile. As will be seen below, there are two types of day matching, neither of which has formally been shown to have the statistical property of unbiased estimation.

3.1.1 Prior day averaging

For use in estimating individual non-residential customer baseline loads,[10] Goldberg and Agnew (2003) recommend using the average load profile of days prior to the event scaled to match the event day's morning loads, thereby partially correcting for weather effects (equations given in the Appendix). The benefits of this approach include easy implementation and intuitive results.

For residential program evaluation, however, prior day averaging with scaling is troublesome for many reasons. For example, it is designed to evaluate the response during a single event, so evaluation of an entire season would require more complex analysis. Also, where programs or tariffs provide for advance notification of events, a method that calculates baselines using the day before and adjustments based on load a few hours before the event are likely to be biased. This can lead to overestimation of response where pre-cooling strategies are employed or underestimation of response where load drop precedes the exact event onset. Load shifting effects can be calculated using a prior-day baseline that excludes any hours after notification is given; however, these calculations must be done separately for each day, making the analysis less transparent. Finally, this method does not control for any potential day-of-week effects and might not accurately control for weather effects.

3.1.2 Weather-based matching

Herter, McAuliffe and Rosenfeld (2006) identify baseline loads for a residential pricing pilot by averaging load profiles of non-event days with weather conditions closely matching those of the critical peak pricing (CPP) event days (equations given in the Appendix). A direct comparison of average baseline loads and average observed loads on the CPP-event day yields the average load impact estimates.

While transparent, weather-based matching can be difficult to implement if there are too few non-event days with weather characteristics similar to the event day. For example, if CPP events are always called on the hottest days of the year, there may not be enough hot non-event days to create a statistically believable baseline. Even if one can find a non-event day with identical weather as the CPP-event day, these two days are not exactly matched due to the chronological differences that can affect baseline loads. Finally, as in

[10] Goldberg and Agnew (2003) do not comment on the potential usefulness of this method for residential program evaluation.

the prior-days averaging technique, estimating the spill-over effects of a CPP event can only be done separately, one day at a time, doubling or tripling the effort needed to find the effects on the event day alone.

3.2 Regression techniques

3.2.1 Regression-based load profile comparison

A regression based-load profile comparison is a statistical implementation of the prior day averaging and weather matching techniques based on analysis of covariance (ANCOVA) in studying hourly loads (model specification given in the Appendix). This method can be used to calculate response to either a dynamic rate or a load control program (see for example Quantum, 2006).

One can estimate a single model for all load data across customers (e.g., Hartway, Price and Woo, 1999; Aigner and Hirschberg, 1985; Aigner and Lillard, 1984).[11] However, we consider customer-specific regressions to be advantageous for several reasons. First, the method is flexible in that the customer-specific response estimates can be averaged to any desired degree of detail (e.g., local weather zone vs. state-wide), or even used individually if desired. This is an important consideration in program design, evaluation and implementation. A program that induces large response in a given area but not others might be better suited for a targeted rather than state-wide implementation.

Second, it eliminates the need for customer demographic data, because variation of demographic variables within-customer over a short time period (e.g., one or two summer seasons) is unlikely to be large. This is not the case for regression models that pool all customer data, because demographic characteristics vary significantly across customers. This is an important benefit of individual versus pooled regressions for program evaluation, because demographic characteristics are unlikely to be available for all customers in the sample.[12]

This regression-based approach offers a number of advantages. First, its empirical implementation is relatively straightforward, using the least squares regression routine available in any statistical package. For example, using SAS, the estimation of a large

[11] With hourly dummy variables, a single model may be one hourly load regression for all customers, which can be accurately and easily estimated using ordinary least squares (OLS). Alternatively, a single model can be a system of 24 hourly equations, each of which corresponds to a particular hour. The system can be jointly estimated using the seemingly unrelated regression technique, possibly with cross-equation constraints to improve statistical efficiency.

[12] For large-scale programs, utilities are likely to have only hourly load and weather data for each customer, since a survey of participant demographics is a costly and time-consuming proposition, and missing data can bias results. A TAG member comments that if the research focus is average response, the missing demographic data may not be an issue. The comment is valid if the probability of a variable (e.g., income) with missing values does not correlate with the size of kW load. If customers who do not report income are also large users, excluding these customers from the regression can cause sample selection bias (Heckman, 1979).

number of customer-specific regressions can be easily done on a personal computer using the BY option of PROC REG.[13]

Second, the approach is a statistical implementation of the intuitively appealing day-matching techniques, isolating the customer-specific kW impact of a CPP event while controlling for both time (as in prior-day averaging) and weather (as in weather-matching). This technique further improves on the day-matching techniques by easily calculating spillover effects on the days preceding and following an event. Thus, the approach can accurately quantify heterogeneous customer responses that may be weather-dependent, without the laborious exercise of finding matching days.

Third, the customer-specific response estimates found via least squares have the desirable statistical properties of being unbiased and consistent (Kmenta, 1986).[14] An unbiased estimate is accurate: any deviation between the estimate and its true but unobserved value has an expected value of zero. A consistent estimate converges to its true but unobserved value when the sample size increases. Hence, the kW response estimate's validity improves with increasing sample size, which can be the result of a longer sample period during which more data are collected. This is an important consideration for assessing the load impact of an on-going program already in place because updating the load impact estimates after receiving more data over time will move the estimates closer to their true but unobserved values.

Finally, the results of the regression can easily produce intuitive results in the form of average hourly response estimates and load profiles.[15]

3.2.2 Econometric demand analysis

An econometric demand analysis relies on the microeconomic theory of an electricity consumer that can be a household or a firm (see the Appendix for the model specification). It requires sufficient price variations for the identification and estimation of a demand model's price coefficients. As a result, the approach is better suited for a

---

[13] As noted by a TAG member, hourly load data likely has serial correlation. In the presence of serial correlation, OLS produces unbiased, though inefficient, estimates. To improve statistical efficiency, however, one can use autoregression techniques that account for serially-correlated errors.

[14] A TAG member comments that the load-regression approach may fail to produce accurate load impact estimates when the load data only contains as few as one event. Accurate estimates can still be obtained in this special case using the following steps: (1) estimate the baseline load profile after excluding data on the day-before, day-of, and day-after the events; (2) predict the baseline load profile for the day-before, day-of, and day-after the single event; and (3) compare the predicted baseline load profile with the actually metered load data for the day-before, day-of, and day-after the single event.

[15] As noted by a TAG member, a customer-specific regression approach may not yield accurate load impact estimates for the special case that an area has mostly very cool weather and extremely rare hot days on which the utility invokes the high-price or load curtailment signal. This is because the extrapolation of the baseline profile inferred from the data collected on cool days can be an inaccurate representation of the baseline profile on the extremely hot days. To improve accuracy in this special case, one should use a regression that employs a pooled data sample. While plausible, this special case is unlikely. For example, this case does not exist in the critical peak pricing data sample that geographically encompasses the entire California (Herter, McAuliffe and Rosenfeld, 2006).

pricing program/experiment than a load control program/experiment, as the latter is seldom designed for an investigation of price-induced effects on a customer's load profile.

In the case of a household, the approach postulates a system of electricity demand equations which, if one chooses, can be derived from the household's utility-maximization behavior (Varian, 1984, Chapter 3).[16] Estimating the electricity demand system under time-of-use (TOU) pricing yields price elasticity estimates by TOU period (Aigner, 1986; Acton, 1982). Notable examples are Atkinson (1979), Aigner and Hausman (1980), Caves Christensen and Herriges (1984, 1987), and Herriges, Caves and Christensen (1984).[17] Faruqui and George (2005) apply a similar framework for their CPP analysis.

In the case of a firm, the approach assumes a system of electricity demand equations derived from the firm's cost-minimization behavior (Varian, 1984, Chapter 1). Estimating the electricity input demand system under TOU pricing yields price elasticity, as demonstrated by Hirschberg and Aigner (1983), Woo (1985), Tishler (1983) and Taylor, Schwartz and Cochell (2005).

Implementation of an econometric demand framework to analyze customer response to high prices is not simple. Specifying and estimating a system of 24 hourly demand equations with numerous parameters is an extremely complicated exercise (Taylor, Schwartz and Cochell, 2005). One can reduce the number of parameters to be estimated by grouping load data in broad TOU periods (e.g., on- and off-peak), but the grouping implicitly treats within period loads identically. Response estimates obtained under this type of restricted specification then represent only average load values across the CPP event hours, and do not reflect any hourly shape.

Second, the approach is not easily amenable to analyze the day-before and day-after spill-over effects in customer response. To estimate these effects, a suitably specified demand system would allow inter-day substitution of hourly kW, expanding the already large number of parameters to be estimated. This worsens the non-convergence problem in estimating a system of hourly demand equations (Zarnikau, Landreth, Hallett and Kumbhakar, 2006).[18]

Third, the numerous elasticity estimates do not clearly convey the kW magnitude of demand response; this is notwithstanding that the kW estimates are of most interest to resource planners and policy makers. While one may use price elasticity estimates to make load impact predictions (Faruqui and George, 2005),[19] a direct and transparent

---

[16] As noted by a TAG member, a demand model may be purely driven by price/quantity data, without making the assumption of utility maximization.
[17] While one can postulate a kW demand system as shown in the Appendix, the primary focus of these TOU studies is price elasticity estimation (Aigner, 1986).
[18] The estimable form of a demand model is not always linear. For instance, the Generalized Leontief (GL) expenditure function implies non-linear cost share equations to be estimated using iterative techniques (Woo, 1985). When the number of equations is large and the sample size is also large, non-convergence is a frequent problem that is not easy to overcome.
[19] There are many definitions of price elasticity (e.g., elasticity of substitution, own- and cross-price elasticities of demand based on electricity expenditure, and own- and cross-price elasticities of demand

method is preferable, especially when the parametric specification of a demand system can at times wrongly influence the size of elasticity estimates (Woo, 1985).[20]

Finally, an econometric demand model may not be applicable to a direct-load-control (DLC) program that allows an electric utility to interrupt/curtail a customer's loads. This is partly because a DLC program seldom produces sufficient price variations for the identification and estimation of the demand model. Assuming sufficient price variations, one may consider using econometric models of household/firm behavior to estimate the effect of service interruption on consumption (Woo, 1994; Woo and Lo, 1993). However, extending that approach to a system of hourly electrical demand equations is difficult because of the resulting large number of parameters to be estimated.

3.3 Recommendation

The key finding from the preceding discussions is that the regression-based load comparison is a reasonable approach to accurately estimate demand response to dispatchable events. This finding leads to our recommendation of using this approach for residential demand response estimation based on load data collected from an experiment or on-going program.

4. CRITICAL KNOWLEDGE GAPS

Notwithstanding our recommendation, critical knowledge gaps remain. By no means exhaustive, the list of gaps include: (a) validation of the accuracy of response estimates; (b) modeling the effect of consecutively triggered events on response estimates; (c) modeling the effect of a day-ahead notice in the presence of public appeal for voluntary load curtailment; and (d) estimating aggregate load impact under a voluntary tariff. Each gap is discussed below.

4.1 Validation of method accuracy

Recall that demand response is the difference between (a) the unperturbed load profile absent the event; and (b) the perturbed load profile as a result of the event. Since the perturbed profile is directly and accurately metered, a response estimate can only be as accurate as the unperturbed baseline profile estimate.

Unfortunately, the unperturbed load profile can never be directly observed, unless one can meter a customer's unperturbed loads under the exact conditions surrounding the event day. This is an impossible task because the customer does not have an identical

---

based on total household expenditure that includes electricity) (Acton, 1982). While these elasticity values are algebraically related (Berndt and Wood, 1979), how to use them correctly is not immediately obvious to someone with limited exposure to electricity demand studies.

[20] A TAG member comments that a linear demand system can be used directly to predict baseline loads and load impact directly, without using elasticity estimates. This comment is valid in the context of using an *estimated* version of the linear demand system. However, it presupposes the system has already been estimated. As shown in the Appendix, a linear demand system is neither simple nor easy to compute when compared to the other alternatives.

twin with the same consumption behavior living next door in an identical house. As a result, the unperturbed load profile can at best be a prediction, using a model developed from data collected from the non-event days.

While direct and accurate metering of demand response is impossible, validating the accuracy of response estimates can be done using the following steps:

- Formally establish that the baseline load estimation is unbiased. For instance, an hourly load regression estimated using least squares has unbiased coefficient estimates that can in turn produce unbiased load predictions.
- Empirically verify that the within-sample load predictions match the unperturbed and perturbed load profiles in the sample used to build the load estimation model.[21]
- Empirically verify that the beyond-sample load predictions match the unperturbed and perturbed load profiles outside the sample used to build the model.[22]

If an approach is found to be unbiased and able to produce accurate within- and beyond-sample load predictions for both event and non-event days, we can conclude that the approach's accuracy is validated to the best extent allowed by the available data.

4.2 Further requirements for accurate program evaluation

Consecutive events

How consecutive events affect demand response is seldom studied, even though hourly load data files (e.g., CPP experiment) are available for estimating this effect. The effect is important when the utility must invoke load reduction to resolve a persistent emergency (e.g., record 1-week heat wave plus major plant failures). If a customer's demand response declines after the first event day, the program's value diminishes as well.

To quantify the effect of consecutive events on load impact, a customer-specific regression can include a binary indicator, as an additional explanatory variable, to reflect if an event-day is immediately preceded by another event day, see the appendix for further details.

Advance notice

The CPP data indicate that advance notice induces demand response on the day of notice, even though the high CPP price is not in effect. Separately, past experience from the 2000/2001 energy crisis indicates that customers respond to public appeal for voluntary reduction in an emergency declared by the California Independent System Operator (CAISO). A substantive question thus arises: if the advance notice coincides with the

[21] A TAG member comments that the average of OLS within-sample predictions always matches the average of within-sample actual load data. While the comment is valid, one should check if the daily predictions also match well with the actual data on days of interest (e.g., very hot non-event days and event days).

[22] An example of beyond-sample prediction comparison entails the following steps: (1) estimate the load regressions using the first two months of a 3-month sample period; (2) use the estimated regression equation to predict the kW in the last month; and (3) compare the actual and predicted kW in the last month.

emergency, what should be the kW effect of the advance notice? The answer is important, if one were to correctly attribute the value of advance notice.

Aggregate load impact

A program's cost effectiveness critically depends on the value of its aggregate impact, the product of (a) the average impact per participant; and (b) the number of participants (= participation rate x number of eligible customers). The estimation methods discussed thus far aim to provide estimates for the average impact in (a) due to a voluntary tariff / program.[23] It does not discuss how (a) can be estimated under mandatory participation. Nor does it estimate the participation rate required to compute the number of participants in (b) under voluntary participation.

While mandatory implementation implies 100% participation, possible self-selection bias (Ham, Mountain and Chan, 1997) complicates the use of the data of a voluntary experiment to infer the average load impact. Self-selection bias arises because a participant of a voluntary program reveals itself to behave differently from a non-participant. It can be identified and corrected using the procedure developed by Heckman (1979).

A demand response tariff can be voluntary, with an "opt-in" or "opt-out" implementation. Hence, the load response estimation discussed in Section 3 can produce an average load impact per participant. But a voluntary tariff's participation rate is below 100% because an eligible customer who dislikes the tariff can switch to another one. To quantify the tariff's participation rate, one may apply discrete choice modeling that explains the probability of a customer's participating decision (Hartman, Doane and Woo, 1991; Keane, MacDonald and Woo, 1988).

5. SUMMARY

A good demand response estimation method should meet the criteria of simple implementation, accuracy, and usefulness:

- Is the method easy to implement? A method is undesirable if it requires extensive training requirement and is time-consuming to apply.
- Are the results accurate? Under- or over-estimating a program's kW savings leads to under- or over-statement of the program's demand response value. Hence, accuracy is an overarching goal of any demand response estimation method.
- Are the results useful? Transparency facilitates third-party review and validation. A black-box approach is undesirable because it invites skepticism, diminishing a

[23] As a TAG member noted, if the estimates are based on the data collected in an experiment or the initial year of a small program, one may have concern in extrapolating the estimates to the case of wide implementation because of the potential bias caused by the small sample size. This concern, however, diminishes as estimates can be updated using an increasingly large sample resulting from the program's on-going operation.

demand response program's acceptance by various stakeholders (e.g., ratepayers, utilities and regulators).

Based on a review of several methods for evaluating demand response, we find that customer-specific regression analysis is likely to give accurate, transparent and intuitive results. Depending on program requirements, this method can be modified to estimate hourly demand response before, during and after events, providing hourly kW response results and load profiles.

This recommendation is purely based on our conceptual reasoning and a careful review of estimation literature. No empirical results have shown this method to be easier to implement, more accurate or to provide more useful results than the other methods described herein. Such empirical evidence is critical to a final recommendation.

Beyond basic demand response estimation, several issues need to be addressed before a practical method for residential demand response program evaluation can be determined. Among them are the ability to evaluate multiple events on consecutive days, an understanding of how advance notification affects demand response, and incorporation of considerations affecting the extrapolation of results from a voluntary pilot to a large-scale program.

6. REFERENCES

Acton, J.P. (1982) "An Evaluation of Economists' Influence on Electric Utility Rate Reforms," *American Economic Review* 77(2): 114-119.

Aigner D.J. (1986) "The Residential Electricity Time-of-Use Experiments: What Have We Learned?" in J.A. Hausman and D.A. Wise (editors) *Social Experimentation*, NBER University of Chicago Press, IL: Chicago.

Aigner, D.J. and Hausman, J.A. (1980) "Correcting for Truncation Bias in the Analysis of Experiments in Time-of-Day Pricing of Electricity," *Bell Journal of Economics* 11(1): 131-142.

Aigner D.J. and Hirshberg J.G. (1985) "Commercial/Industrial Customer Response to Time-of-Use Electricity Prices: Some Experimental Results," *Rand Journal of Economics* 16(3): 341-355.

Aigner D.J. and L.A. Lillard (1984) "Measuring Peak Load Pricing Response from Experimental Data: An Exploratory Analysis," *Journal of Business and Economics Statistics* 2(1): 21-39.

Atkinson, S. (1979) "Responsiveness to Time-of-Day Electricity Pricing: First Empirical Results," *Journal of Econometrics* 9: 79-95.

Baskette, C. Horii, B., Kollman, E. and Price S. (2006) "Avoided Cost Estimation and Post-Reform Funding Allocation for California's Energy Efficiency Programs" *Energy* 31: 1084-109.

Berndt, E.R. and D.O. Wood (1979) "Engineering and Economic Interpretation of Energy-Capital Complementarity," *American Economic Review* 69(3): 342-354.

Caves. D., Christensen, L. and Herriges, J. (1984) "Consistency of Residential Customer Response in Time-of-Use Electricity Experiments," *Journal of Econometrics* 26: 179-203.

Caves, D., Christensen, L. and Herriges, J. (1987) "The Neoclassical Model of Consumer Demand with Identically Priced Commodities: An Application to Time-of-Use Electricity Pricing," *Rand Journal of Economics* 18 (4): 564-580.

Faruqui, A. and George S. (2005) "Quantifying Customer Response to Dynamic Pricing," *The Electricity Journal* 18(4): 53-63.

Goldberg, M.L. and Agnew, G.K. (2003) *Protocol Development for Demand Response Calculation - Findings and Recommendations*, 400-02-017F. Sacramento, CA: California Energy Commission.

Ham, J., Mountain, D. and Chan, M. (1997) "Time-of-Use Prices and Electricity Demand: Allowing for Selection Bias in Experimental Data," *Rand Journal of Economics* 28(1): 113-141.

Hartman, R.S., Doane M.J. and Woo C.K. (1991) "Consumer Rationality and the Status Quo," *Quarterly Journal of Economics*, February, 141-162.

Hartway, R., Price, S. and Woo, C.K (1999) "Smart meter, Customer Choice and Profitable Time-of-Use Rate Option," *Energy* 24: 895-903.

Heckman J. (1979) "Sample Selection Bias as a Specification Error," *Econometrica* 47: 153-161.

Herriges, J., Caves. D. and Christensen, L. (1984) "Modeling Alternative Residential Peak-Load Electricity Rate Structures," *Journal of Econometrics* 26: 249-268.

Herter, K., McAuliffe P. and Rosenfeld A. (2006) "An Exploratory Analysis of Customer Response to the California Critical Peak Pricing Pilot," *Energy* forthcoming.

Hausman, J.A. (1981) "Exact Consumer Surplus and Deadweight Loss," *American Economic Review* 71: 662-676.

Hirschberg, J. and Aigner, D. (1983) "An Analysis of Commercial and Industrial Customer Response to Time-of-Use Rates," *Energy Journal* 4: 103-126.

Horowitz, I. and Woo, C.K. (2006) “Designing Pareto-Superior Demand-Response Rate Options,” *Energy* 31: 1040-1051.

Keane, D.M., MacDonald, L.S. and Woo, C.K. (1988) "Estimating Residential Partial Outage Costs with Market Research Data," Special Electricity Reliability Issue, *Energy Journal*, 9: 151-159.

Kmenta J. (1986) *Elements of Econometrics*, MacMillan, NY: New York.

Mood, A.M., Graybill, F.A. and Boes D.C. (1974) *Introduction to the Theory of Statistics*, McGraw Hill, NY: New York.

Quantum Consulting Inc. and Summit Blue Consulting, LLC. (2006) *Evaluation of 2005 Statewide Large Non-residential Day-ahead and Reliability Demand Response Programs*, *Final Report*, Prepared for Southern California Edison Company and Working Group 2 Measurement and Evaluation Committee. Berkeley, CA.

Taylor, T., Schwarz, P.M. and Cochell, JE (2005) "24/7 Hourly Response to Electricity Real-Time Pricing with up to Eight Summers of Experience," *Journal of Regulatory Economics* 27(9): 235-262.

Tishler, A. (1983) "The Industrial and Commercial Demand for Electricity Under Time-of-Use Pricing," *Journal of Econometrics* 23: 369-384.

Varian, H. (1984) *Microeconomic Analysis* Norton, NY: New York.

Woo, C.K. (1985) "Demand for Electricity of Small Nonresidential Customers under Time-of-Use (TOU) Pricing," *Energy Journal* 6(4): 115-127.

Woo, C.K., Hanser, P. and Toyama N. (1986) "Estimating Hourly Electric Load with Generalized Least Squares Procedures," *Energy Journal* 7(2): 153-170.

Woo, C.K. (1994) "Managing Water Supply Shortage: Interruption vs. Pricing," *Journal of Public Economics*, 54: 145-160.

Woo, C.K. and Lo K.W.K. (1993) "Factor Supply Interruption, Welfare Loss and Shortage Management," *Resource and Energy Economics* 15: 339-352.

Zarnikau J., Landreth G., Hallett I. and Kumbhakar S.C. (2006) *Industrial Customer Response to Wholesale Prices in the Restructured Texas Electricity Market*, Frontier Associates, TX: Austin.

7. APPENDIX: MODEL DETAILS

7.1 Prior-days averaging

The unadjusted baseline for hour j is:

$$Baseline_j = \frac{\sum_{n=1}^{N} Load_{nj}}{N}$$

where $Load_{nj}$ is the load on day n in hour j and N is the number of days used to calculate the baseline.

A scalar adjustment for weather can then be calculated as:

$$AdjBaseline_j = Baseline_j + (Load_{ex} - Baseline_x)$$

where $Load_{ex}$ is the load on the event day in hour x, the target hour for adjustment (typically one to three hours before the event). Finally, response is calculated as:

$AdjBaseline_j - Load_{ej}$

7.2 Weather-matching technique

This technique quantifies the average hourly response in a given temperature ranges as:

$$\text{Response}_{ij} = \frac{\sum_{p=1}^{P_{ij}} \left(\frac{\sum_{e=1}^{E_{ijp}} Load_e}{E_{ijp}} \right)}{P_{ij}} - \frac{\sum_{p=1}^{P_{ij}} \left(\frac{\sum_{n=1}^{N_{ijp}} Load_n}{N_{ijp}} \right)}{P_{ij}}$$

where: i = temperature range; j = hour of the day j (1-24); p = participant; e = event day; n = normal day; P_{ij} = number of participants having both normal and event values in temperature range i for hour j; E_{ip} = number of event days in temp range i for participant p; N_{ip} = number of normal days in temp range i for participant p; $Load_e$ = kWh usage for participant p, temperature bin i, and hour j on event days; $Load_n$ = kWh usage for participant p, temperature bin i, and hour j on normal days.

7.3 Regression-based load profile comparison

Load regression

Consider a customer-specific hourly load regression with coefficient vectors (β, α_1, α_2, α_3):

y_{th} = $\boldsymbol{x}_{th}\,\beta + D_{1t}\,\mathbf{z}_{th}\,\alpha_1 + D_{2t}\,\mathbf{z}_{th}\,\alpha_2 + D_{3t}\,\mathbf{z}_{th}\,\alpha_3 + \varepsilon_{th}$; or

Hourly kW = Hourly baseline kW for hour *h* on day *t*

+ Day-before ΔkW impact (e.g.., advance notice effect if *t* = day before the event)

+ Day-of ΔkW impact (e.g., intra-day shifting if *t* = day of the event)

+ Day-after ΔkW impact (e.g., rebound effect if *t* = day after the event)

+ Random error.

The above regression is to be estimated for each customer with the following variables:

y_{th} = observed kW load of a participant in hour *h* on day *t*.

$\boldsymbol{x}_{th}$ = (1 x *K*) vector of explanatory variables that drive the kW load on a "normal" day, include the intercept, kW timing and weather variables.

$\mathbf{z}_{th}$ = (1 x *M*) vector of variables that drive load impact, with $M = K + 2$. It has all of the elements in $\boldsymbol{x}_{th}$. The first additional variable is the total number of events in the season as of day *t*, so as to control for the possible effect of event operation history on customer response. If an increase in the cumulative number of events reduces load impact, this variable's coefficient estimate is negative. The second is a binary variable, indicating if the event day is immediately preceded by another event-day. This variable aims to capture the effect of consecutively triggered events on the event day's load impact. If consecutive events reduce load impact, this variable's coefficient estimate is negative.

D_{1t} = 1 if day-before an event; 0, otherwise.

D_{2t} = 1 if day-of an event; 0, otherwise.

D_{3t} = 1 if day-after an event; 0, otherwise.

ε_{th} = error term, assumed to have zero mean and finite variance.

Based on Kmenta (1986, Chapter 11), Table A.1 shows that equation (1) can yield estimates of the hour's baseline load, unperturbed by an event. It can also isolate the load change as a result of the event.

Table A.1: Load type, dummy variable values and load expectation

Load type	Dummy variable values	Expected value	Explanation
Day-before baseline	$D_{1(t-1)} = 0$ $D_{2(t-1)} = 0$ $D_{3(t-1)} = 0$	$\boldsymbol{x}_{(t-1)h}\ \beta$	$\boldsymbol{x}_{(t-1)h}\ \beta$ is the normal load that could have occurred based on the observed values of $\boldsymbol{x}$ on the day before the CPP event.
Day-before load impact	$D_{1(t-1)} = 1$ $D_{2(t-1)} = 0$ $D_{3(t-1)} = 0$	$\boldsymbol{z}_{(t-1)h}\ \alpha_1$	$\boldsymbol{z}_{(t-1)h}\ \alpha_1$ is the load impact based on the observed values of $\boldsymbol{z}$ on the day before the CPP event.
Day-of baseline	$D_{1t} = 0$ $D_{2t} = 0$ $D_{3t} = 0$	$\boldsymbol{x}_{th}\ \beta$	$\boldsymbol{x}_{th}\ \beta$ is the normal load that could have occurred based on the observed values of $\boldsymbol{x}$ on the day of the CPP event.
Day-of load impact	$D_{1t} = 0$ $D_{2t} = 1$ $D_{3t} = 0$	$\boldsymbol{z}_{th}\ \alpha_2$	$\boldsymbol{z}_{th}\ \alpha_2$ is the load impact based on the observed values of $\boldsymbol{z}$ on the day of the CPP event.
Day-after baseline	$D_{1(t+1)} = 0$ $D_{2(t+1)} = 0$ $D_{3(t+1)} = 0$	$\boldsymbol{x}_{(t+1)h}\ \beta$	$\boldsymbol{x}_{(t+1)h}\ \beta$ is the normal load that could have occurred based on the observed values of $\boldsymbol{x}$ on the day after the CPP event.
Day-after load impact	$D_{1(t+1)} = 0$ $D_{2(t+1)} = 0$ $D_{3(t+1)} = 1$	$\boldsymbol{z}_{(t+1)h}\ \alpha_3$	$\boldsymbol{z}_{(t+1)h}\ \alpha_3$ is the load impact based on the observed values of $\boldsymbol{z}$ on the day after the CPP event.

Prediction

Suppose we have used least squares to consistently estimate equation (1), obtaining ($\boldsymbol{b}$, $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$), the estimates for (β, α_1, α_2, α_3). Based on Kmenta (1986, Chapter 10), the discussion below is specific to the day-of predictions. The formulae for day-before and day-after predictions are entirely analogous.

Suppose we are interested in the predicting the day-of baseline load at hour h on day T with a scenario described by $\boldsymbol{x}_{Th}$. We can make $\boldsymbol{x}_{Th}$ to reflect a very hot day when the system demand spikes. The day-of baseline load prediction is:

$$y_{Th} = \boldsymbol{x}_{Th}\, \boldsymbol{b}. \tag{2}$$

Conditional on a scenario described by $\boldsymbol{z}_{Th}$ (whose value should be consistent with $\boldsymbol{x}_{Th}$), the day-of load impact prediction is:

$$\Delta_{2Th} = \boldsymbol{z}_{Th}\, \boldsymbol{a}_2. \tag{3}$$

Since the load impact is customer-specific and based on the customer's own load data observed over event and non-event days, it does not suffer from self-selection bias.

Average of customer-specific predictions

The hourly baseline load prediction based on equation (2) is customer-specific; so is the hourly load impact prediction based on equation (3). Averaging is necessary if our interest is the average load impact for a particular customer stratum (e.g., single family owners in a hot weather zone). Suppose this stratum has N customers. After suppressing the day and hour subscript for notational simplicity, the stratum's average load impact is:

$$\Delta \quad = \quad \Sigma_j \, \Delta(j) / N \qquad (4)$$

where $\Delta(j)$ = load impact estimate for customer j (= 1, ..., N).

The variance of the average load impact is (a) the variance of individual customer-specific response estimates,[24] divided by (b) the number of customers in the segment (Mood, Graybill and Boes, 1974).

We can use the stratum-specific average impact and its variance to compute the system average and its variance. As the sample's proportion of customers in a given stratum may differ from the population's proportion, we apply inverse weights (= sample proportion / population proportion) to inflate or deflate the influence of each stratum's estimate in the system average computation.

7.4. Econometric demand analysis

Hourly demand equation

Consider a linear hourly demand equation for hour h' on day t:

$$q_{h't} \quad = \quad \theta_{h'} + \Sigma_h \, \gamma_h \, p_{ht} + \Sigma_h \, \phi_h \, p_{h(t+1)} + \Sigma_j \, \eta_{h'j} \, w_{h'j} + \Sigma_k \, \delta_k \, d_k + \mu_{h't}$$

where $q_{h't}$ = kW at hour h' on day t (h' = 1, ..,, 24); p_{ht} = applicable price at hour h (= 1, ..., 24) on day t; $p_{h(t+1)}$ = applicable price at hour h on day (t+1); $w_{h'tj}$ = weather variable j for hour h' on day t; and d_k = demographic variable k (e.g., income, house size, and appliance holding) that do not vary within the relatively short sample period (e.g., one summer); and μ_{ht} = error term. The coefficients to be estimated are: θ_h = hourly specific intercept, $\{\gamma_h\}$ = coefficients for the day-of prices, $\{\phi_h\}$ = coefficients for the day-ahead prices; $\{\eta_{h'j}\}$ = coefficients for the weather variables, and $\{\delta_k\}$ = coefficients for the customer demographic variables.

This demand equation measures day-of and day-after price effects using the coefficient vectors $\{\gamma_h\}$ and $\{\phi_h\}$, weather effects $\{\eta_{h'j}\}$, and demographic effects $\{\delta_k\}$. It captures the notice effect because the prices on the notice day t are less than those on the event day

[24] As noted by a TAG member, the individual variance estimates should reflect the estimation method chosen (e.g., PROC REG vs. PROC AUTOREG in SAS).

(t + 1). When day t and (t+1) are both non-event days, the prices in these two days are identical. If the price effects are assumed to depend on weather and demographics, the price coefficients can written as linear functions of those variables.

Since each day is represented by 24 hourly equations per day, there are 48 hourly demand equations to be estimated using such methods as the iterative three-stage least squares or maximum likelihood. Even without allowing for price effects' dependence on other variables, the number of parameters to be estimated per equation can easily exceed 50, in light of the 48 hourly prices, several weather variables, and several demographic variables.

Once can reduce the number of price coefficients by grouping hours by time-of-use period. For example, an on- and off-peak grouping would reduce the number of price variables and the number of equations to four, at the expense of imposing the assumption of equal kW effects of price, weather and demographic within each period.

Prediction

To the extent that the system of equations is correctly specified, the resulting coefficient estimates are unbiased and can be used to make accurate baseline predictions under the assumption of no CPP price signal. The same set of estimated equations can be used to make perturbed load predictions under the assumption of CPP price signal being dispatched. The hourly difference between two load predictions forms the hourly demand response estimate.

Average of customer-specific predictions

This is the same as the one for regression-based load profile comparison; hence, it is not repeated here.