DR Impact Evaluation – Which Design and Analysis Method Is Right for What?

Craig Williamson, Applied Energy Group, Denver, CO
Kelly Marrin, Applied Energy Group, Riverside, CA

ABSTRACT

Unlike Energy Efficiency (EE) impact evaluations, which usually focus on energy savings, Demand Response (DR) impact evaluations focus on hourly load impacts, particularly but not exclusively during events and at the time of the system peak. Like EE evaluation, estimating impacts requires careful planning up front and the choice of an appropriate analysis method. There are several analysis methods appropriate to estimate DR impacts, including difference of differences, regression, and price elasticity models. Each has advantages and disadvantages; none is appropriate in all cases. Difference methods are simpler and more direct, require fewer assumptions, and are easier to understand. But they are not as flexible or adaptable. Regression models allow for more adjustment, calibration, and interpretation, but can have issues with model assumptions. Elasticity studies add additional assumptions and structure to the model, which may not always apply, but allow for a better understanding of price response. Control group issues can arise across all these methods, both with randomized control trials and quasi-experimental designs, and must be handled differently for pilots versus fully implemented programs. In this paper we discuss the critical importance of study design up front, and then discuss the strengths and weaknesses of the different analysis methods.

Design and analysis methods really do matter. To get valid, accurate, and unbiased estimates of DR impacts, those managing and evaluating DR programs need to use the best method for their program, goals, and circumstance.

Introduction

Whether the program is part of a pilot or an implemented program, the effective evaluation of Demand Response (DR) programs requires a thoughtful evaluation plan that ensures the analysis is appropriate for both the pilot design (if applicable) and the program itself. In this paper, we first discuss design and planning considerations for evaluation of pilots and programs. We then describe analysis options, including the strengths and weaknesses of the various methods that can be used for estimating impacts.

Throughout this paper we focus primarily on Demand Response, including direct load control (AC and water heater switches and thermostat setbacks) and event-based dynamic pricing programs. These programs can also result in energy savings, for which we often need impact estimates as well, but this paper will not discuss the design and analysis of energy efficiency (EE) programs.

Planning and Design – Always Critical

Before embarking on the evaluation of any pilot or program, the first step is to define the goals. For a pilot, these will usually be research questions that need answers. The more specific the questions, the better. For a program, the goal is often a reduction in peak demand, but this, too, can be more narrowly defined. Many efforts have failed because the goals are not clearly defined up front.

It is important to consider how program or pilot impacts will be estimated from the very beginning. We discuss the considerations for pilots and programs separately here, looking at pilots first, and then at programs. Pilots are efforts primarily intended to test the feasibility of a program, work out details of implementation, study customer preferences, and estimate impacts for a potential future roll out. Usually the goal of estimating savings and load reductions from pilots is for understanding and prediction, not to claim impacts in a regulatory proceeding or offset specific generation during the pilot. Programs, on the other hand, are implemented to get results. A DR program is put in place to reduce demand. Some DR programs, particularly dynamic pricing programs, result in overall energy savings as well as demand reductions, but the primary focus is on demand reduction.

These differences between pilots and programs mean that the planning and evaluation of pilots and programs also differ.

Pilots

Before embarking on programs that involve untested technologies, new pricing structures, or innovative approaches to demand reduction, it is prudent to conduct a pilot. A pilot can be carefully designed and controlled, allowing for unbiased estimates of demand reductions, as well as for testing different implementation options. Pilots can be used to test new technologies, in terms of both functionality and load reductions. For instance, a pilot could test different cycling strategies and thermostat setbacks for air-conditioning (AC) demand response programs. Pilots have also been used to test different dynamic pricing schemes. To reiterate, the goal of these pilots was not to maximize demand reductions during the pilot, but to understand how different factors affect the impacts.

Randomized Control Trial Designs. Because the focus is not on achieving demand reductions, we have more freedom to design and control pilots. The traditional "gold standard" for experimental design is a randomized control trial (RCT). The concept of an RCT design is simple—customers are randomly assigned to treatment and control groups, and the program is only implemented for those in the treatment group. This random assignment (called randomization) ensures that before the start of the pilot, the two groups are equivalent in all ways, both measurable and unmeasurable. The randomization, if properly done, ensures there is no bias in the assignment or in the subsequent estimate of the impacts. All customer characteristics will be approximately equally represented in both groups. Because of this, any differences between the treatment and control groups during the treatment period (after the implementation of the program) will be due to the program, with a measurable level of statistical uncertainty. The level of statistical uncertainty will depend on the sample size as well as on the inherent variability of the savings (usually energy use or demand).

An important aspect of an RCT design is that customers must be assigned to treatment and control groups through the randomization process, without allowing any customers to choose which group they are in, or leave the study based on their assignment. Self-selection bias is often an issue and can manifest in different ways. Self-selection bias results from the fact that customers who choose to participate in a program are inherently different from those who choose not to participate, and also from those who are not given that choice. If customers who choose to participate in a program are put into a treatment group, and those who do not choose to participate are put in the control group, that would result in self-selection bias – there would be no way to know if the differences were due to the program or due to pre-existing differences between people who choose to participate and those who do not. Customers must be assigned to treatment and control based on the randomization, and not influenced by any customer choices.

Most traditional load control DR programs are voluntary. Pricing programs can be voluntary, but can also be offered on a default basis, usually with an opt-out provision. While intuitively RCT designs may seem more applicable to a default, opt-out environment, they can also be used for voluntary programs. For voluntary offerings, however, all customers, both in the treatment and control group, must first volunteer for the program and then be randomly assigned to either treatment or control. The key point is that the consistent random assignment to the treatment or control group from the same population of customers using an RCT, regardless of whether participation in the program is voluntary or mandatory, eliminates self-selection bias in the estimation of impacts.

Randomizing customers for a default pilot is straightforward – customers are selected to participate in the pilot, and then are randomly assigned before there is any communication with the customers about the program. Each customer has an equal chance of being a participant or control customer. Care must be taken to ensure there is no program effect that "bleeds" into the control group, and no marketing or other programs are directed at the control group unless they are also directed at the treatment group.

Randomization for a voluntary pilot is more difficult. The key is to randomize customers so treatment and control groups consist only of customers who have tried to sign up for the program. This can be done by putting a process in place where customers that sign up are randomly assigned to treatment or control during signup. Those assigned to the control group are told either that they do not qualify and can't participate in the program (referred to as a "recruit and deny" approach) or that they can sign up but not until sometime later, often the following year (called a "recruit and delay" approach). The most direct method of doing this is to randomize "on the fly," where customers are assigned dynamically during signup. It is also possible to "pre-randomize" the entire population and create a database with indicators for treatment and control. When a customer signs up, the database is checked and, based on their original assignment, the customer is told either that they can sign up or that they cannot sign up. Both methods, if carefully implemented, will create a properly randomized and unbiased control group. This pre-randomized approach with a recruit and delay was used for the successful Smart Study TOGETHER pricing pilot at Oklahoma Gas and Electric (OG&E) in 2010-11.

It may seem like this approach for a voluntary program would introduce self-selection bias, since customers are choosing to participate. The key here is that both treatment and control groups are made up of customers who want to participate, but those assigned to the control group are not allowed to participate. This means the impacts estimated based on the difference between the groups will reflect only the effect of the program and be unbiased, since differences between those who choose to participate and those who don't will not affect the estimates. However, the results of the study will only apply to those who choose to participate. This is most appropriate for a pilot where the future program is expected to be voluntary, with all customers allowed to choose whether to participate or not. The impacts would not reflect the savings for a full roll-out of a mandatory program to all customers. Some refer to applying the results of a voluntary pilot to a mandatory future program as introducing self-selection bias; that bias is in the inappropriate application of the results not in the impact estimate.

The drawback with the approach described above for a voluntary program is that customers who want to sign up but are not allowed to will often be disappointed. This can lead to customer dissatisfaction, especially if not done carefully. Steps can be taken to mitigate customer dissatisfaction, such as offering alternate compensation (that must be unrelated to energy consumption) or clarifying up front that not all who want to sign up will qualify.

Quasi-Experimental Design. When it is not practical or acceptable to recruit and deny or recruit and delay, an alternative to the RCT Design is a quasi-experimental design with a matched control group. A matched control group can be created after the participants have been recruited for a voluntary program. The control group is developed by selecting control customers that are as similar as possible to each treatment customer during the pre-treatment period, based on known observable characteristics. This can result in a control group well matched on observable characteristics, but, unfortunately, we can never know whether or not it is well matched based on unobservable characteristics. The most important unobservable characteristics are those related to what drove the customer to volunteer for the pilot in the first place, since any control group will inherently exclude customers who wanted to sign up. However, when an RCT design is not possible, a matched control group is usually the best alternate approach, with the highest likelihood of giving less biased results.

There are a couple of approaches to creating a matched control group. Generally, a control group pool of non-participating customers is created, and one non-participant customer is matched to each participating customer based on some measure of similarity. These methods also usually define filters or groupings to ensure the control group pool for each participating customer includes only customers that are similar. Examples of filters are geography (i.e. look only at customers in the same zip code as the participant), central AC ownership, demographic segmentation, and the like.

One approach of choosing the best match within the control group pool for each participant is to select customers by minimizing the Euclidean difference (the square root of the sum of squared deviations) between the participant and control customers across as many characteristics from the pre-treatment period as possible. Another is propensity score matching, which uses a model to identify what pre-treatment characteristics drive customers' propensity to sign up, collapsing those characteristics into a propensity score, and then matching on similar propensity scores. Both methods require a large pool of potential control group non-participants from which to select the matches, but there are some situations where one or the other of the two performs better.

Once all the matching control customers are determined, they can be analyzed the same way a randomized control group is analyzed. One advantage of a matched control group is that the participant population can be subdivided, with each participating customer pulling along their match into their subgroup, without introducing any bias. This is not possible for an RCT if the subdividing criteria are based on something about the participating customers (such as those who opt out of events, for instance).

Within-Subjects Design. Another option is to use each customer's own energy use as a control. This is known as a Within-Subjects Design. For DR pilots, this can take two forms. We can compare the customers' loads before the start of the program with their loads during the treatment period. Used by itself, the impact estimates for this design can be affected by differences across time, including weather, economics, and even technology. If the treatment period is an exceptionally mild summer, and the pre-treatment period is a very hot summer, the load reduction will be significantly overestimated if that difference is not taken into account. If the economy is improving, customers may purchase more appliances that use energy, and increase their energy use, or could eat out more often, thereby decreasing energy use. Because of these issues, within-subjects designs are less effective when used across years.

However, certain DR programs allow for another type of within-subjects design. During a pilot days without events can be used as a control group for days with events. It is important to take this into consideration when calling events since events should not be called on all extreme days. A protocol can be developed that involves predicting whether the next day will be an extreme day, and if it is, an alternating pattern (call every other day as an event) can be used to ensure a pool of comparable non-event days.

One important consideration is that a within-subjects design based on non-event days will only work well if the DR program has no effect on non-event days. This is particularly effective for AC Cycling or thermostat setback programs where the customers' AC units are controlled by the utility and only on event days. In this situation days on which events are not called are a reasonable representation of what the customers' loads would be if the program were not in place. In contrast, an event-based Critical Peak Pricing (CPP) program combined with a time-of-use (TOU) rate, where customers pay a different on-peak and off-peak price on all weekdays and then pay a much higher price during events, would not work for this approach. Customers would likely change their energy use patterns on all days in response to the TOU prices so comparing event days with non-event days would give a biased savings estimate. To reiterate, this approach of using non-event days as a control for event days is only appropriate for DR programs that are not expected to affect energy use on non-event days.

Programs

When programs are being implemented or rolled out to a large segment (or all) of the population, many of the same considerations apply but the situation is different. All three methods described above can be used, but present different challenges for programs being implemented.

The first major difference is that the goal of an implemented DR program is to reduce the demand at the time of the system peak as much as possible; because of this choosing not to call events on all extreme days is usually not an option. Furthermore, holding out half the population for a control group is impractical. We now discuss ways to adapt the above methods to fully implemented programs.

Randomized Control Trial Designs. The RCT design is the least adaptable to a full program roll-out. RCT designs are used for some programs, most notably large roll-outs of behavioral programs such as Opower's Home Energy Reports. These are considered programs, with utilities claiming savings, and even can include a DR component. However, utilities taking this approach are finding their customer base is quickly "used up" by the requirements for control groups. One way to alleviate this problem somewhat is to use smaller control groups – say a control group of only 20,000 customers for a treatment group of 100,000 or more. As long as the assignment is randomized using a smaller control group is fine and creates no bias but it can result in less precise estimates of savings, which can be problematic for behavioral programs with small average impacts. Beyond this, RCT designs are generally not used when programs are implemented.

Quasi-Experimental Design. Matched control groups, on the other hand, can work reasonably well for programs as they are rolled out. The weaknesses described above remain—we can never know what unobservable differences between participants and non-participants that may affect impact estimates.

The same considerations for creating the matched control group apply here though it can be challenging to find a big enough control group pool from which to select the matches. For this reason, this tends to work better for programs that represent a relatively small proportion of the total population. The potential unobserved differences are inversely proportional to the amount of marketing done. With minimal marketing and recruiting, many potential control group customers may not even be aware of the program, and so may not differ as much from the participants as those who are aware, but choose not to participate. In one particular dynamic pricing evaluation, the consensus decision was not to use a matched control group, since the program had been heavily marketed for several years, with a significant percentage of the available population participating in the program. Those customers not participating had refused repeatedly, and were thought to be quite different from those that decided to participate. For this evaluation, we used a Variation in Adoption approach (described below),

Within-Subjects Design. This approach is most applicable to programs that have been rolled out; it does have the same potential issues but may be the only option available. Including both the customer's pre-treatment and treatment period energy use in the analysis can work if as many other factors are included in the model as possible. One twist on this is the Variation in Adoption (VIA) approach. If customers are joining a program across a wide time frame, and the recruiting, marketing, and targeting do not change appreciably during that time, future participants can be used as a "built-in" control group. Because they are future participants, they should be more similar to the current participants than any non-participants would be.

For those DR programs which only affect energy use on event days, just as with pilots, using non-event days as a control group for event days can work very well. However, since the goal of the program (and the basis for its cost-justification) is to reduce load at the time of the system peak, not calling events on the most extreme days can be problematic. There are two ways around this. One is to model the energy use across the hours of the day, including temperature variables, based only on non-event days, and then use the model to predict the baseline that the DR customers would have used if an event had not been called. This can work, but it is challenging when there are few or no extreme non-event days, since the model has nothingto base the relationship between energy use and temperature on for those extreme temperatures. Such an approach is only practicable when the predictions are made for days with temperatures that fall within the range of the historical (modeling) data or are only marginally outside that range. Researchers should explicitly note whenever the model is used to predict outside of the range of historical temperature (or other explanatory variables).

One other possible adaptation of the within-subjects design for broadly implemented DR programs is to randomly leave out a small subset of customers from each of (or some of) the events. These uninterrupted customers can serve as a baseline for the remainder of the participant population. The validity of this "temporary control group" can be checked by examining other non-event days and comparing them to the rest of the population on those days. This group could even be selected taking this into account. If the entire participating population is split into numerous random subgroups—all of which are representative of the whole—the customers left out can be varied across events resulting in a more "fair" administration of the program. While this adaptation may not be possible in all cases, if the control technology is capable of targeting different subsets of customer it is an attractive option.

Analysis Methods

Once a pilot or program is in place, and has been operating for a period, it is time to analyze the data and produce results. Generally speaking, for DR impact estimation, there are three approaches: difference in differences (DID), regression, and elasticity modeling. Not all analysis methods can be used with all designs, but there are options with each.

DID is the most direct approach and requires fewer assumptions but is the least flexible. A regression approach is more flexible and allows for weather and other factors to be incorporated into the savings estimates. This additional flexibility enables the estimation of both ex post and ex ante savings estimates, but requires some assumptions about the nature of the relationships between variables. Elasticity models provide estimates of different price elasticities, which allow for interpretation of price response more broadly, and can be used to adjust impact estimates across a wide range of prices. However, elasticity estimates also require more rigorous analysis design and more assumptions than a basic regression model. We now discuss each of these in detail.

Difference in Differences

The aptly-named Difference in Differences (DID) method involves taking the difference between the control group and treatment group energy use during both the treatment period and the pre-treatment period, and then subtracting the pre-treatment difference from the treatment period difference. The differences are done at the group level, based on the average across all customers. The simplest model would be based on annual energy, with only four average values involved, one pre-treatment and one treatment period value for the treatment group and the same for the control group. Equation 1 shows the calculation of the estimate of savings using a difference in differences, and holds for energy use data in any time frame (annual, monthly, daily, or hourly).

$$Savings = (X_1 - Y_1) - (X_0 - Y_0) \qquad (1)$$

where X is the control group, Y is the treatment group, subscript 0 refers to the pre-treatment period and subscript 1 refers to the post-treatment period. Using algebra, this can be rewritten as an adjusted baseline (the control group, less the pre-treatment difference) minus the treatment group, shown in Equation 2 below. The advantage of this form is that it compares the actual treatment customer energy use and a baseline, and allows for the calculation of savings as a percent of that baseline.

$$Savings = [X_1 - (X_0 - Y_0)] - Y_1 \qquad (2)$$

Using basic statistics, standard errors can be calculated and used to determine statistical significance and confidence intervals on the savings. This standard error can also be used to perform a hypothesis test to see if the savings are statistically significant. The significance and hypothesis test formulas require the variance of the savings as an input. The variance of the savings depends on: the variance of the energy use for treatment and control, the covariance of energy use between the pre-treatment period and the treatment period, and the sample sizes.

The equations above reflect a direct calculation of the DID estimate. It assumes the treatment and control group are independent and randomized. The test of significance and the estimation of the confidence intervals assume the sample is large enough that the data follow the Central Limit Theorem. Most pilots and programs include hundreds or thousands of customers so that is well above the 30 customers usually assumed to be the minimum needed. This direct calculation requires no other assumptions on the nature of the relationships between energy use and other variables. This lack of other assumptions and the fact that it is unbiased makes it an attractive method.

An alternate to this direct calculation is to use a regression model to estimate the DID savings. If correctly formulated, the DID regression will provide the same savings estimates and the same significance and confidence intervals as the direct calculation. However, care must be taken to ensure all the model uncertainty is taken into account in the calculation of the standard error of the estimates.

The main benefit of using a DID regression formulation over the direct calculation is that it can be used to account for the impacts of other factors on overall energy use and on savings. The most common form this takes is the impact of temperature, usually expressed as heating and cooling degree days. The presence of different forms of cooling or electric heat can also be included in the model. With this regression formulation, the more other factors you include that explain variation in energy use, the more precisely the model can estimate changes in energy use due to the program.

The additional benefits of using a regression formulation come with a cost, which is the need for additional data, and the associated assumptions. These assumptions are described in detail in the regression section below.

An effective way to fix both the problem of correlation of errors and the difficulty of accounting for a substantial portion of energy use in the model is to use a fixed effect regression model. In the next section, we describe the fixed effects model, and go into more depth on considerations and issues that come along with using regression models, many of which also apply to the DID regression model as well.

- *The following are strengths of DID analysis*: flexibility; simple and direct calculation and estimation; results in unbiased estimates; able to adjust for pre-treatment differences; very few assumptions needed,
- *The following are weaknesses of DID analysis*: does not allow for scenarios with weather or other factors; correctly accounting for covariance can be challenging; cannot account for other factors (without a DID regression).
- *The DID approach is appropriate for*: pilots or programs; RCT; matched control group; within subjects design; ex post impact estimates.

Regression

Many types of regression models can analyze energy use data to estimate load reduction for DR programs and pilots. The most common is the fixed effects regression model. The idea underlying a fixed effect regression model is fairly simple. Most regression models include an intercept term, which in our case represents the base energy use that does not vary based on the other factors in the model. A fixed effect regression model estimates a separate intercept term for each customer. This fixed effect can account for many customer-specific characteristics that influence energy use without having to gather survey or audit data. It also accounts for pre-treatment differences in average usage between the treatment and control group, since each customer (including treatment and control customers) has its own base level of energy use, and the savings vary from that base.

Equation 3 below shows an example of a fixed effects regression model that estimates the load reduction on event days. The model would be based on data for all summer weekdays, with the dependent variable being the energy use from, say, 2-5 pm, the timeframe for events being called.

$$kwh_{it} = \alpha_i + \beta_1 CDD_{it} + \beta_3 Event_t + \beta_4 (Event_t * CDD_{it}) + \varepsilon_{it} \qquad (3)$$

Where the variables and their coefficients are defined as:

kwh_{it} = consumption of customer *i* on day *t*
α_i = a fixed effect for each customer *i*
$\beta_1 CDD_{it}$ = the Cooling Degree Days (CDD) for customer *i* on day *t*
$\beta_3 Event_t$ = an indicator variable that takes on a value of 1 if *t* is an event day, and 0 otherwise
$\beta_4 (Event_t * CDD_{it})$ = an interaction term between the event day indicator variable and the CDD for customer *i* on day *t*
ε_{it} = the error for participant *i* on day *t*

In this simple model, the on-peak energy is a function of the customer-specific intercept, the temperature (expressed as cooling degree days or CDD), and whether or not the day is an event day. The load reduction on a particular event day could be estimated as the coefficient on the event day indicator plus the number of CDD for that day times the coefficient on the event day-CDD interaction variable.

This model would be appropriate for a direct load control program or pilot under a within-subjects design where there is no change in energy use expected on non-event days. It would require that there were high temperature non-event days in the datasets to enable the model to estimate the coefficients correctly. For a design with a subgroup of customers excluded from each event, the event day variable would not be set to 1 for those customers who did not receive an event signal on a given event day. If the pilot was a randomized control trial, a coefficient for participation could replace the event day variable, with that variable set only for the participants.

While this is probably simpler than the final model used for most DR evaluations, it illustrates the method well. The model is developed based on the data and the coefficients of the model can be used to estimate the impacts on any given day. With CDD in the model impacts can be estimated for future days based on possible temperatures, even if those specific temperatures did not occur during events in the past. However, care should always be taken in extrapolating results of a regression model outside the domain used to estimate the model. For instance, a model based on event days that ranged from a high temperature of 90 to 98 degrees will not accurately estimate the load reduction on a day with a high of 105 degrees.

Regression models require certain assumptions in order to be valid. The standard assumptions are that the relationships between the independent and dependent variables are linear (or can be transformed to be linear), and that the errors (residuals) are independent and identically distributed (IID) following a normal distribution with mean zero and constant variance. The assumption of linearity is not difficult to ensure, but care must be taken, especially around models that include CDD. The relationship between energy and CDD is often linear for small segments of the domain but not necessarily for the entire range of temperatures. Because of this we often create multiple degree day coefficients based on different temperature levels, requiring only that the relationship is linear for each segment (referred to as piecewise linear).

The requirement for the residuals to be IID normal can be more challenging. Without a fixed effect for each customer, the residuals for each observation (day or hour) for a customer are highly correlated—bigger customers have all larger positive residuals and smaller customers all larger negative residuals. The fixed effect model helps to solve this issue. There can also be autocorrelation, with the residuals on successive observations correlated, but this can be dealt with by using robust standard errors.

- *Regression strengths*: relatively flexible; able to account for pre-treatment differences using a fixed effect or indicator variable; easily able to incorporate weather; useful for scenario analysis.

- *Regression weaknesses*: imposes some assumptions on the data; requires a more expertise and experience to execute well.
- *Regression is appropriate for*: pilots or programs; RCT; matched control groups; within subjects design; ex-post, ex-ante, and weather normalized impacts.

Elasticity Models

For DR programs based on time-varying prices, an elasticity model can be used to estimate impacts and to predict impacts based on pricing scenarios outside of those specifically implemented in the pilot. Price elasticity is generally defined as the percentage change in the quantity of a good or service demanded in response to a percentage change in price. We now describe elasticity modeling as applied to DR based on time-varying prices. We provide more background and context here, since readers may be less familiar with elasticity modeling and price elasticity.

Elasticity estimates can be used to estimate changes in energy consumption given changes in price. These estimates are especially useful for predicting rate impacts under different weather and pricing scenarios. The impacts estimated using hourly or monthly models as described above allow us to measure the actual demand response achieved during a program year, and allow us to adjust that demand based on different (i.e., normal) weather. The elasticity estimates allow us to predict the amount of demand response we can expect under different pricing scenarios. Elasticity estimates are useful to predict how customers will respond to changes in price; they are much less practical for straightforward estimation of impacts.

For elasticity estimates to be valid, the model should conform to one of the generally used functional forms used to model consumer demand. Based on a review of method used to estimate elasticities in 7 time-differentiated rate pilot impact evaluations, three basic models are being used.

- The first is a simple linear **double log model**. A double log model is a simple and flexible modeling framework that can be easily estimated using Ordinary Least Squares (OLS) regression. However, because it is not based firmly in economic consumer theory it can sometimes be considered ad-hoc.
- The second option is a **Constant Elasticity of Substitution (CES)** demand system. The CES system usually employs two equations. The first equation estimates the elasticity of substitution. The elasticity of substitution measures the ease with which one good, off-peak energy can be substituted for another good, on-peak energy. The second equation is used to estimate the daily elasticity, which estimates participants' response to average daily prices. The two equations together can be used to estimate the impact during the on-peak period and the daily change in consumption. The CES framework is relatively simple to estimate using either OLS or Generalized Least Squares (GLS), is firmly based in consumer demand theory, and is widely used within the industry.
- The third option is a **Generalized Leontief (GL)** form. The main advantage to using a GL framework is that it relaxes the assumption made by the CES form that the elasticity remains constant. It is most commonly used to estimate shifting during shoulder periods in price structures with more than one on-peak period, where price response might be different during different on-peak periods. For example, the shifting in the morning shoulder might be substantially different than the shifting in the afternoon shoulder.

In general applications, with relatively simple price structures, the CES demand system is usually the best choice. It can be used to estimate both an elasticity of substitution, and a daily elasticity. The CES demand system specification is flexible, relatively simple to estimate, and is industry best practice based on the literature review.

Elasticity of Substitution Equation. The elasticity of substitution equation models the ratio of on-peak to off-peak kWh usage as a function of the ratio of on-peak to off-peak prices. In many cases, when there is technology such as a programmable communicating thermostat (PCT) being tested, the

PCT can be included in the model. The researcher must consider factors other than prices, temperature, and PCT, not accounted for in the model, that may lead to time-of-day shifting by consumers. While prices should be viewed as the best signal, if behavior is affected due to other factors (media, other EE programs)elasticity estimates may be biased. Accounting for other factors may not be practicable in the model but could/should be explicitly acknowledged in the researchers. An example of an Elasticity of Substitution model is presented below in Equation 4.

$$ln\left(\frac{kWh_p}{kWh_{op}}\right)_{it} = \alpha_i + \beta_1\, ln\left(\frac{p_p}{p_{op}}\right)_{it} + \beta_2\left(CDH_p - CDH_{op}\right)_{it} + \beta_3 PCT_{it} + \beta_4 Seas_t + \beta_5 ln\left(\frac{p_p}{p_{op}}\right)_{it}\left(CDH_p - CDH_{op}\right)_{it} + \beta_6 ln\left(\frac{p_p}{p_{op}}\right)_{it} PCT_{it} + \varepsilon_{it} \tag{4}$$

Where the variables and their coefficients are defined as:

$ln\left(\frac{kWh_p}{kWh_{op}}\right)_{it}$ = the natural logarithm of the ratio of on-peak kWh to off-peak kWh for customer i on day t

α_i = a fixed effect for each customer *i*

$\beta_1\, ln\left(\frac{p_p}{p_{op}}\right)_{it}$ = the natural logarithm of the ratio of on-peak price to off-peak price for customer i on day t

$\beta_2\left(CDH_p - CDH_{op}\right)_{it}$ = the difference between the on-peak and off-peak Cooling Degree Hours (CDH) for customer i in time t

$\beta_3 PCT_{it}$= an indicator variable that takes on a value of 1 for customers with a PCT

$\beta_4 Seas_t$= seasonal indicator variables i.e. month, year, and day of week

$\beta_5 ln\left(\frac{p_p}{p_{op}}\right)_{it}\left(CDH_p - CDH_{op}\right)_{it}$= the interaction between the price ratio and the difference between on and off-peak CDH

$\beta_6 ln\left(\frac{p_p}{p_{op}}\right)_{it} PCT_{it}$= the interaction between the price ratio and PCT

ε_{it} = the error term for customer i on day t

The model specified in Equation 4 allows us to estimate the elasticity of substitution and also quantify how the elasticity of substitution changes under different weather conditions, and with the addition of the PCT. The elasticity of substitution for the participants is the sum of the price terms as shown in Equation 5 below.

$$ES = \beta_1 + \beta_5 * \left(CDH_p - CDH_{op}\right) + \beta_6\, PCT \tag{5}$$

Where:

ES = the elasticity of substitution

β_1 = the effect of changes in price alone

$\beta_5 * (CDH_p - CDH_{op})$ = the incremental effect of weather

$\beta_6\ PCT$ = the incremental effect of the PCT

Daily Elasticity Equation. The daily elasticity equation models average daily kWh usage as a function of the average daily price. As above, the model can capture changes in the daily energy based on changes in weather and the presence of a PCT. An example of a Daily Elasticity model is presented below in Equation 6.

$$ln(kWh_{it}) = \alpha_i + \beta_1\, lnP_{it} + \beta_2 CDH_{it} + \beta_3 PCT_{it} + \beta_4 Seas_t + \beta_5 ln(P_{it})CDH_{it} + \beta_6 ln(P_{it})PCT_{it} + \varepsilon_{it} \tag{6}$$

Where the variables and their coefficients are defined as:

$ln(kWh_{it})$= the natural logarithm of average daily kWh for customer i in time t

α_i = a fixed effect for each customer *i*

$\beta_1\ lnP_{it}$= the natural logarithm of the average daily price for customer i in time t

$\beta_2 CDH_{it}$ = the average cooling degree hours (CDH) for customer i in time t

$\beta_3 PCT_{it}$= an indicator variable that takes on a value of 1 for customers with a PCT

$\beta_4 Seas_t$= seasonal indicator variables i.e. month, year, and day of week

$\beta_5 ln(P_{it})CDH_{it}$= the interaction between the price and CDH

$\beta_6 lnP_{it}PCT_{it}$= the interaction between the price and PCT

ε_{it} = the error term for customer i in time t

The model specified in Equation 6 allows us to estimate the daily elasticity and also quantify how the daily elasticity changes under different weather conditions, and with the addition of the PCT. The daily elasticity for the participants is the sum of the price terms as shown in Equation 7 below.

$$ED = \beta_1 + \beta_5 * CDH + \beta_6\ PCT \tag{7}$$

Where:

ED = the daily elasticity

β_1 = the effect of changes in price alone

$\beta_5 * CDH$ = the incremental effect of weather

$\beta_6\ PCT$ = the incremental effect of the PCT

Once the elasticities have been calculated the combination of the two can be used to predict rate impacts. Keep in mind the main focus of an elasticity model is the elasticity itself, and because the depended and independent variables represent rates of change rather than unit estimates, the interpretation of the coefficients and the estimation of savings is not nearly as straightforward as with a typical regression model developed for evaluation.

- *Elasticity model strengths*: ability to predict rate impacts under different pricing and weather scenarios (the only method of the three that can do this); results include elasticity estimates.
- *Elasticity model weaknesses*: much more complex execution and estimation; many assumptions made about the nature of the relationship of price to energy use; interpretation of results is more complex; not practical for general evaluation purposes.
- *Appropriate for*: studies with the goal of estimating price response; pilots only, with RCT, matched control group, or within subjects design.

References

George, S., Jimenez, L., Potter, J. 2014. *SmartPricing Options Final Evaluation.* Prepared by SMUD for the U.S. Department of Energy. September 5.

Global Energy Partners. 2011. *Evaluation of OG&E's SmartStudy TOGETHER Pricing Pilot, Final Report.*

Charles Rivers Associates. 2005. *Impact Evaluation of the California Statewide Pricing Pilot*. March 16.

Marrin, K., Williamson, C. 2013. "Matched Control Group Methods: What's Your Propensity?" presentation to the Western Load Research Association Spring 2013 Conference, http://www.wlra.net/recentpapers.htm.

Nguyen, Abigail. 2014. "The Often Ignored Covariance of Two Random Variables," presentation to the Western Load Research Association Spring 2014 Conference, http://www.wlra.net/recentpapers.htm.

Summit Blue Consulting LLC. 2006. *Evaluation of the 2005 Energy-Smart Pricing Plan Final Report*. prepared for Community Energy Cooperative.

Summit Blue Consulting LLC. 2007. *Final Report for the MyPower Pricing Segment Evaluation,* December 21.

The Brattle Group. 2009. *BG&E's Smart Energy Pricing Pilot Summer 2008 Impact Evaluation*. April 29.