Title
Quantifying the Impacts of Timebased Rates, Enabling Technology, and Other Treatments in Consumer Behavior Studies: Protocols and Guidelines

Author
Cappers, Peter

Publication Date
2013-06-20

LBNL-XXXX

ERNEST ORLANDO LAWRENCE
BERKELEY NATIONAL LABORATORY

Quantifying the Impacts of Time-based Rates, Enabling Technology, and Other Treatments in Consumer Behavior Studies: Protocols and Guidelines

Peter Cappers, Annika Todd, Michael Perry, Bernie Neenan, and Richard Boisvert

Environmental Energy Technologies Division

June 2013

This work described in this report was funded by the Office of Electricity Delivery and Energy Reliability, under Contract No. DE-AC02-05CH11231.

Disclaimer

LBNL-XXXX

Quantifying the Impacts of Time-based Rates, Enabling Technology, and Other Treatments in Consumer Behavior Studies: Protocols and Guidelines

Prepared for the
Office of Electricity Delivery and Energy Reliability
U.S. Department of Energy

Principal Authors

Peter Cappers, Annika Todd, Michael Perry, Bernie Neenan, and Richard Boisvert

Ernest Orlando Lawrence Berkeley National Laboratory
1 Cyclotron Road, MS 90R4000
Berkeley CA 94720-8136

June 2013

This work described in this report was funded by the Department of Energy Office of Energy Efficiency and Renewable Energy and Office of Electricity Delivery and Energy Reliability, under Contract No. DE-AC02-05CH11231.

Acknowledgements

This work described in this report was funded by the Department of Energy Office of Electricity Delivery and Energy Reliability, under Contract No. DE-AC02-05CH11231 and the Electric Power Research Institute (EPRI).

The authors would like to thank Joe Paladino (DOE OE) for his support of this project.

This publication is a corporate document that should be cited in the literature in the following manner:
Quantifying the Impacts of Consumer Behavioral Study Experiments and Pilots: Protocols and Guidelines. LBNL, Berkeley, CA and EPRI, Palo Alto, CA: 2013. LBNL-XXXXX.

Table of Contents

List of Figures and Tables

Acronyms and Abbreviations

AMI	Advanced Metering Infrastructure
ARC	Aggregator of Retail Customers
AS	Ancillary Service
BA	Balancing Authority
CAISO	California Independent System Operator
DLC	Direct Load Control
DR	Demand Response
ERCOT	Electricity Reliability Council of Texas
FERC	Federal Energy Regulatory Commission
I/C	Interruptible and Curtailable
ILR	Interruptible Load for Reliability
ISO	Independent System Operator
ISO-NE	Independent System Operator New England
IRC	ISO/RTO Council
IRP	Integrated Resource Planning
IOU	Investor-Owned Utility
M$/yr	Million dollars per year
MCP	Market Clearing Price
MISO	Midwest Independent System Operator
MW-h	Megawatt per hour
NERC	North American Electric Reliability Corporation
NJBPU	New Jersey Board of Public Utilities
NPCC	Northeast Power Coordinating Council
PJM	PJM Interconnection, LLC
PSCo	Public Service of Colorado
RPS	Renewable Portfolio Standard
RTO	Regional Transmission Organization
SPP	Southwest Power Pool
U.S.	United States

Executive Summary

This report offers guidelines and protocols for measuring the effects of time-based rates, enabling technology, and various other treatments on customers' levels and patterns of electricity usage. Although the focus is on evaluating consumer behavior studies (CBS) that involve field trials and pilots, the methods can be extended to assessing the large-scale programs that may follow. CBSs are undertaken to resolve uncertainties and ambiguities about how consumers respond to inducements to modify their electricity demand. Those inducements include price structures; feedback and information; and enabling technologies embedded in programs such as: critical peak, time-of use, real-time pricing; peak time rebate or critical peak rebate; home energy reports and in-home displays; and all manner of device controls for appliances and plug loads. Although the focus of this report is on consumer studies—where the subjects are households—the behavioral sciences principles discussed and many of the methods recommended apply equally to studying commercial and industrial customer electricity demand.

The report is written from the perspective of an analyst who evaluates pilots and field trials. It links choices made in the experimental design to analysis methods that are applicable to the design. In other words, the report addresses how best to ascertain whether interventions produced the intended and significant changes in electricity demand. Because experiments and pilots can be and are designed in many different ways, a wide range of methods is discussed. They share the goal of precisely measuring whether changes in electricity usage are caused by the intervention being tested. This guide serves as a starting point to help analysts decide what should be done and understand what it takes to accomplish that end. Extensive references provide the required technical details.

Background

DOE and EPRI share an interest in establishing protocols for analyzing the results of CBS pilots and field experiments. DOE has funded 11 CBS projects as part of its Smart Grid Investment Grant program. EPRI is supporting utilities in fielding CBS project as part of its Smart Grid Demonstration project. In some cases, both DOE and EPRI are involved in the design and evaluation of the studies. In all cases, they share a commitment to ensuring that rigorous scientific practices and protocols are applied so that the results are useful to the project host utility and can be extrapolated to wider circumstances. This ensures that the resources expended to conduct these studies have widespread applicability.

EPRI and DOE, through Lawrence Berkeley National Laboratory (LBNL), collaborated on guidelines for designing experiments for Smart Grid research (EPRI report 1020342, 2010; DOE 2010, and EPRI has issued guidelines on designing CBS studies (EPRI report 1025734, 2012). This report contributes to that body of work by addressing how to evaluate the results of pilots and experiments.

Objectives

This report provides analysts responsible for evaluating CBS pilots and experiments with a single-source primer on the methods and practices available—and when each is applicable. Guidance on choosing a method is provided along with information on their rigorous

applications and full and transparent reporting of the results. This report advances the understanding of how consumers use and value electricity; in particular, it characterizes the ways in which behavioral inducements can be employed to modify electricity demand.

Approach

The project began with a schematic representation of robust experiments that can be designed to define the range of situations an analyst will encounter. Each was associated with a specific set of issues that must be addressed in evaluating the results. These determine the extent to which the analysis is straightforward. For example, analysis of variance sufficiently tests for significant effects, but unanticipated or unavoidable conditions may result in the design failing to comply with the rigorous provisions that define a randomized control trial— the Gold Standard for statistical testing of effects.

Results

The analysis protocols address the specific condition of the experiment, adjusting (as required) for intervening factors that may have undermined what was initially a randomized control trial (RCT) design. Corrective measures are provided for cases in which a RCT is not possible but when the intent is to estimate the population impact of the behavioral interventions.
Applications, Value, and Use

Utility program evaluators can use this report to develop an evaluation plan that produces credible results which will allow others to understand what was done and why. Pilot and experiment designers will find these protocols helpful in choosing a design and anticipating all of the data that will be required to produce useful findings.

1. Introduction

In the past several decades, there has been much study and debate about how customers use and value electricity. Improved understanding of electricity consumption behavior would help the industry to find better ways to achieve energy and peak demand savings. Furthermore, technology advancements—including advanced metering infrastructure (AMI), smart meters, and wireless communications—allow for innovative and less costly ways to encourage efficient patterns of consumption through time-based rates, enabling technology, and feedback.

Research to portray and characterize customer electricity consumption behaviors accelerated about 30 years ago, in large part because of federal and local legislation and regulatory mandates. In the 1980s, issues related to load research, cost allocation, marginal cost estimation, rate design, and time-differentiated pricing were addressed through various research activities. During that time, several utilities implemented time-of-use rates, both mandatory (primarily for large general-service customers) and voluntary (for all customers).

In the 1990s, federal and state governments in some parts of the United States began to restructure and deregulate the U.S. electric power industry to facilitate the creation and management of wholesale power markets and retail competition. As the reforms spread, research programs at the retail level on many of these issues slowed, while existing programs were either capped to new enrollment or dismantled altogether.

Over the past decade, utilities have increasingly sought to install AMI or smart meters. Programs that offered customers inducements to alter their electricity use, either on an ongoing basis (price response) or under limited and specific circumstances (demand response), were generally an integral part of the benefits assessment that justified the investment (described in EPRI report 1017006, 2008). Regulators and stakeholders demanded a greater level of detail about the performance of these programs to justify the utility's purported benefits because prior research efforts had not focused on the types of opportunities enabled by AMI.

The result was a new wave of pilots and experiments designed to quantify the effects of a variety of behavioral inducement programs on participants' electricity use, focused on both the timing and level of those load changes.

Although evaluations of these research efforts share many common methods and protocols, they differ substantially in their execution and level of detail to which methods and results are reported. Some of these differences are a result of the variety of rate treatments (for example, real-time pricing, variable peak pricing, critical peak pricing, and peak time or critical peak rebate) that were evaluated. Analysis methods applicable to time-differentiated rate schedules (for example, time-of-use [TOU]) are not always applicable to rates that vary infrequently (for example, critical peak pricing [CPP]).

In addition, these studies were conducted under a variety of circumstances. Some were initiated by utilities interested in clarifying performance impact levels; others were mandated by regulators in a way that dictated the research agenda as well as the experimental design. Still others were shaped by many stakeholders with different interests. As a result, the pilot program designs and evaluation efforts were often narrowly focused, at the expense of a more general investigation of the impacts hypothesized to be most important. As such, the application of different analytical methods and approaches to the reporting of the results has made it difficult to compare the conclusions drawn across studies.

It is in part because of these combined circumstances that many important aspects of how customers use and value electricity remain unanswered. There are several serious gaps in the understanding of our ability to influence customer demand through time-based rates, enabling technology, feedback, and education (EPRI report 1025856). There are gaps in our knowledge about both the size and persistence of the load impacts; the heterogeneity in load impacts across customer segments; and the extent to which the customer response in those pilots can be extrapolated to other customers, utilities, and regions of the country.

There is a plethora of research efforts ongoing to characterize how behavioral inducements influence electricity consumption. More are likely in the near future.[1] The potential exists for these initiatives to fill many of these remaining gaps in our knowledge and advance our understanding of electricity customer behavior. To that end, a substantial effort to provide guidance on the research design, analysis and coordination across studies may eliminate any serious methodological shortcomings and avoid the squandering of scarce resources that would have resulted in a duplication of effort, missed opportunities and/or misleading findings.

1.1 Purpose of This Report

This report constructs, explains, and rationalizes analysis protocols for measuring the effects of time-based rates, enabling technology, and various other treatments on customers' levels and patterns of electricity usage. In particular, protocols for the three critical phases of pilot programs are outlined and discussed: experimental design, analyses for measuring the observed effects on customer electricity usage, and the reporting of the results.[2] Although the focus is on evaluating pilots and experiments, the methods can be extended to assessing the large-scale programs that may follow to verify the initial impact estimates.

---

[1] For example, there is the U.S. Department of Energy's Smart Grid Investment Grants (SGIG) funded through the American Recovery and Reinvestment Act (ARRA). These matching grants were awarded to several electric utilities to fund 11 pilots focusing on time-based rates, enabling technology, and other treatments. A second set of pilots will be undertaken through EPRI's new program, Understanding the Utility Customer (Program 182), whose purpose is to bring a fresh perspective to understanding the behavior of electricity consumers. This program was commissioned to explore new and innovative ways to communicate with customers and actively engage them in decisions that affect electricity usage.

[2] In preparing this report, it is assumed for the most part that the pilot studies have been completed or are at least underway. Therefore, we recommend a range in protocols that applies to a variety of circumstances in terms of questions addressed, experimental design, data availability, and measured effects. The several guidance documents prepared by DOE's SGIG Technical Advisory Group and listed in the references provide valuable advice for conducting pilot studies to those utilities interested in initiating new customer behavioral studies.

The structure and application of the experimental design determine the reference load that serves as the foundation for all analyses. The reference load, also referred to as the counterfactual load, represents (and is an estimate of) what the usage would have been among treatment group customers had they not been exposed to the treatment. Experimental designs include: randomized controlled trials (RCTs), randomized encouragement designs (REDs), matched control group methods, and within-subjects methods. This reference load and its validation are critical to understanding the reliability of the results from the subsequent empirical analyses. These issues are discussed along with a descriptive analysis of the load and other data collected during the pilot, which are believed to be an essential first step in any program evaluation that requires the analysis of such large amounts of data.

The analyses for measuring the effects of various treatments on customer electricity usage are divided into two broad groups:

1. Those that focus on estimating the effects of treatments after the fact, without imposing any specific behavioral structure
2. Those that use an assumed model of decision-making to estimate underlying behavioral parameters

The first category of analyses is almost entirely statistical in nature: established analytical methods are used to estimate the effect of the treatment and indicate the confidence level in the results. The second group relies on economic theory in addition to statistical methods. Formal models are employed that impose the principle tenets of utility (consumers) or profit (businesses) maximization.

We see these categorically differentiated methods as complements. They are part of a series of analyses recommended in this report to ensure that the findings are statistically robust, informative about the nature of behaviors, and actionable for subsequent program design.
In proposing these analytical methods, considerable discussion is devoted to the reporting of the results of the analyses conducted. Only through transparent and thorough description of the analytical method used can we ensure that the results are readily accessible to reviewers, critics, and potential users so that they understand what was done and consider the implications for their design or evaluation effort.

1.2 Roadmap to the Content of the Report

Our goal is to make the measurement of the effects of various treatments on customer electricity usage accessible and understandable to a broad range of evaluators, utility staff, and policy makers with many different backgrounds. We also strive to provide sufficient explanation of analysis methods so that analysts and those who review their evaluations can ascertain which analyses are appropriate to use in certain circumstances. Toward this goal, the report provides a technical exposition to shed light on the potential biases that are part of the variation in measured values and are crucial for interpreting reported results and comparing results across projects.

The discussions of analysis methods that follow in the next three sections describe the circumstances under which they are important along with their applications. These discussions are followed by a more technical exposition to help analysts understand what data are required and which tools are needed to apply the techniques. This requires some familiarity with statistics and economics as well as comfort with mathematical expressions of logical concepts. Those who desire a more in-depth discussion will find details in the footnotes and the appendices, which are supplemented by references to valuable source materials that provide original derivations and examples of their application.

Section 2 outlines a framework for the evaluation of the consumer behavior studies, which includes recommendations for initial descriptive analysis of the data and procedures for establishing a reference load that serves as the basis of comparison with electricity usage measured during the experiment. Section 3 describes methods for conducting statistical analyses to measure treatment effects in a model-free context. Section 4 presents methods for measuring response and impacts using economic models that impose assumptions about the basic tenets of consumer behavior. Section 5 provides a summary of findings.

1.3 Additional Analyses

It is important to emphasize that the discussion that follows does not include one important step in a complete, thorough, and exhaustive examination of the treatment impacts.

A thorough analysis of an experiment would characterize not only what participants did, but also who participated and why. These issues are important for both an opt-in (volunteer subjects) and opt-out (conscripted subjects) design. This understanding of the factors that determine the decision to opt-in or opt-out is important for extrapolating the derived treatment effects to the larger population of customers. If the decision to participate can be associated with observable or measurable customer characteristics, the quantification of the relative importance of those factors and how they correlate with customer responsiveness will greatly improve the effectiveness of recruitment campaigns.

How to conduct this choice modeling is not addressed in this report because it involves a foundational theory and analytical methods that are different from those that define statistical impact analyses and demand modeling. Choice models are explicit expressions of an underlying theory of how consumers make choices.[3] To do justice to the different approaches available

[3] Formal choice models are the ideal vehicle for characterizing participation. They can be used to test hypotheses in which the feature being observed is not measured continuously, such as with energy usage or hourly prices, but rather as a state or conditioned outcome. For example, either individual customers *opted out* of the pilot or they did not—a dichotomous outcome. Another application is to identify causal drivers to observed treatment responses. For example, one may use the subset of participants who are identified as responders to identify what customer attributes and circumstances appear to have contributed to that behavior.

Logit models are regression-based models that are functionally similar to commonly used Ordinary Least Squares (OLS) regression models. However, they differ from other regression models in that they account explicitly for the fact that the outcome is the result of a dichotomous choice. For this reason, the left-hand-side (dependent) variable in the model takes on only values of one or zero, depending on whether the customer chooses to take some action or not (for example, yes/no, buy/not buy, or respond/not respond). The right-hand-side (explanatory) variables are customer characteristics (for example, conventional central air vs. heat pump) and descriptions of the treatments (that is, rebate level) to which the customer has been exposed.

would equally require an exposition of theory and practice much like the one applied in this document, and include conceptual and notational complexity that requires time to absorb. As such, we have chosen to keep this report to a manageable size by focusing exclusively on the analyses for measuring the effects of various treatments on customer electricity usage measurement. We anticipate issuing a stand-alone report that provides a thorough, but understandable, discussion of choice modeling and its application to consumer behavior studies (CBS) analyses.

The estimation of choice models requires information about individual customer circumstances that may be drivers of behavioral change, for example, the stock of household appliances, household income, the number and age of residents, and premise characteristics (such as square footage and efficiency investments).

For an excellent and complete discussion of the logit model and other models of discrete choices, see W. Greene, *Econometric Analysis*. 5th edition, Englewood Cliffs, NJ: Prentice Hall, Inc., 2003, Chapter 21.

2. Framework for the Evaluation of Consumer Behavior Studies

This report defines and describes a set of analyses that, in the evaluation of consumer behavior studies (CBS), address the following fundamental objective:

- Identify how the choices on the levels and patterns of customers' energy consumption are affected by the incentives and information embedded in time-based rates, enabling technology, feedback strategies, and other treatments.

The metrics that address this fundamental question include the following: [4]

1. The average impacts on consumption at specific times of interest (for example, annually, during all events, or during peak periods), which are discussed in Section 3
2. The own-price elasticities of electricity demand and the elasticities of substitution for peak and off-peak electricity demand, which are discussed in Section 4

The analysis of a CBS may extend beyond simply testing for the treatment effect. For purposes of marketing or customer recruitment during program implementation, utilities may also find it important to have knowledge of how socioeconomic characteristics of customers (for example, income or age), premise characteristics (size, appliance stock, and so on), and exogenous factors such as weather affect customers' levels and patterns of usage.

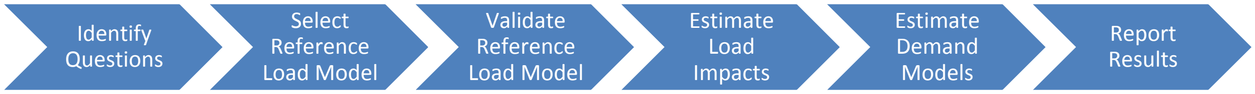
The extent to which these richer characterizations of response in an overall assessment of a CBS can be pursued depends on the experimental design, the objectives of the study, and the data and other information available to address these issues. Several types of data may be available, including energy usage data, survey data on customer demographics and premise characteristics, and data on the interaction between customers and the treatment-related communication and automation/control technology. These data may be available for both the pre- and post-treatment periods, for both control and treatment group customers. The way in which these data are used to enhance the characterization of treatment effects is discussed in Sections 3 and 4.

As noted previously, there is an additional important research objective in the evaluation of any CBS study: characterizing what factors drive treatment participation in instances in which study participants are not conscripted, but subscribed, as is often the case in CBS studies. This understanding can be accomplished through the use of choice modeling analysis methods that associate customer circumstances with the likelihood of agreeing to participate and creates a categorization or segmentation system that informs large-scale program recruitment. Because this objective is antecedent to the establishment of a cause-and-effect relationship for treatments, and because the methods involve an entirely different set of analytical techniques, the discussion of choice models is left to a subsequent inquiry.

[4] Depending on the particular objectives of the study and the data available, DOE's SGIG Technical Advisory Group recommends that some or all of eight impact metrics should be used "to describe the impacts of pricing and information and automation/control technology on customer electricity consumption" (DOE 2010, p.1). The first four metrics are specific forms of the first one listed above and include the average annual electricity impact in kWh, the average hourly impact in kWh, the system coincident peak demand impact in MW, and reliability requirement impacts in MW. The next four metrics are specific forms of the second one listed above and include the own-price elasticity, the cross-price elasticity, the daily price elasticity, and the elasticity of substitution.

2.1 Components of an Evaluation Framework

We propose that there are six essential components to any CBS evaluation as shown in the following graphic. This image is used throughout the report to indicate the topical content of the narrative.



These components are as follows:

- **Identify the specific questions that the analysis must address**. Some of these questions are readily apparent from the types of rates and other treatments in the pilots. For example, an experiment that includes several different rate structures is intended to establish differences among their impacts on electricity usage, perhaps primarily during peak periods or events. The null hypotheses could be constructed as follows: There is no difference in their respective impacts on the level of peak period usage. Likewise, another example might be an experiment that included an in-home display that simply shows loads and another that also allows the consumer to compare current usage to an established goal or a target level dictated by participation in an event-based program. The hypothesis is as follows: There is no difference in the event usage between subjects with the different devices.
- **Other questions might be less obvious but equally important.** For example, recruitment may have been accomplished through distinct engagement steps, and the effectiveness of these steps is an important research question. The hypotheses involve several constituent elements, and the data required to test them are records and process logs. The testing of these (and several dozen more) hypotheses was part of ComEd's Customer Applications Study, an experiment of unprecedented scale and scope (described in EPRI report 1023644).
- **Select the best possible reference load model (or models), given the data available**. The reference load, also known the counterfactual load, is an estimate of what electricity usage would have transpired among treatment group customers had they not been exposed to the treatment. The reference load is determined by the form of the experimental design, the particular questions of interest, and the available data. These issues are discussed in more detail below.
- **Validate the reference load model**. This consists of a demonstration that the model accurately predicts load for the treatment groups under conditions similar to those of interest but where loads are observed. This demonstration of the degree of accuracy is crucial for an outsider to be able to interpret the results. The issue of validation is discussed in more detail below.
- **Estimate load impacts and confidence intervals using the reference load model**. This involves two parts. The first is to establish measurements of the impact based on the reference load and the actual usage data. The second is to subject that measured change (in kWh, for example) to tests of significance. It is not enough that there was a measured change; it must be established that that change was due solely to the treatment. This is accomplished through the calculation of standard errors and confidence intervals and is discussed in Section 3.

- **Estimate consumer demand models needed to derive various price elasticities and elasticities of substitution**. Load impacts are derived by taking the price as given (this seems sensible because it is the effect of the treatment price(s) that we want to evaluate). However, those prices are not necessarily the ones that would be used by the utility in practice. In addition, even if they are, other utilities would likely use different prices—so how can other utilities extrapolate the results of one experiment to its own circumstances? An electricity demand model provides a way to extend the measured price effect to other nominal and relative prices through a metric called price elasticity. These models involve a different structure but use many of the same estimation techniques used in load impact modeling. They are discussed in Section 4.
- **Report the results**. By establishing protocols for reporting research results, we ensure that the experimental design and its validation are well documented, as are the analytical methods used to conduct the empirical analyses. As a result, the estimated load impacts and demand models are communicated in a consistent way that facilitates, to the extent possible, comparability with results across pilots. These reports must be accessible and understandable to a broad range of evaluators, utility staff, and policy makers with many different backgrounds. Recommendations about what information should be documented appear throughout the text where each topic is discussed.

This report is directed toward a broad range of evaluators and utility staff with many different backgrounds. We presume that readers are familiar with basic statistics and econometrics and have the economics background needed for empirical demand modeling. In addition, we assume that readers have basic knowledge of the theory underlying the designs being analyzed. For example, we do not lay out the entire theory behind a randomized controlled trial design (RCT) or a randomized encouragement design (RED). We specify only the particular validation exercises and analyses applicable to each. Finally, we assume that there are certain procedures—such as propensity score matching and bootstrapping—that the reader can learn using standard texts.[5] As stated, however, more details about several of the methods, particularly those related to demand modeling, are found in the footnotes and selected appendices.

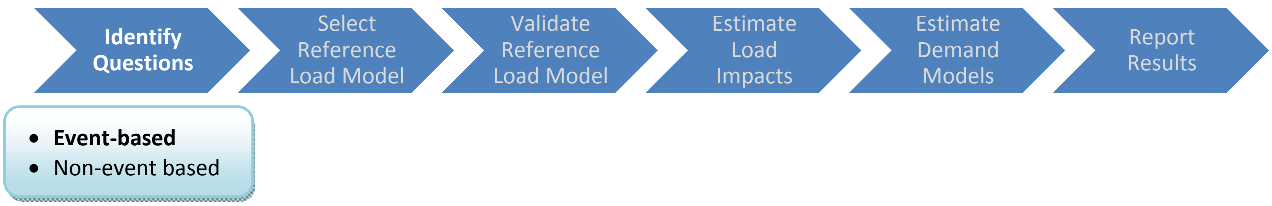
2.2 Characterization of Rate and Other Treatments

New service offerings, which are the subject of experimentation, include highly structured pricing structures as well as various types of enabling technologies and information/feedback mechanisms that may help customers make more informed decisions about electricity use. Enabling technologies, like device or premise controls (for example in-home displays and programmable controllable thermostats) may help customers respond to price inducements so that the combined effect may be greater than what is indicated by the price effect alone. Feedback may also enrich the price response because it gives customers information about when they use electricity. Some have proposed that feedback is a substitute for at least some pricing structures.

---

[5] For a comprehensive book on statistical and econometric techniques, see Greene (2011). For a technical guide to program evaluation, see Imbens and Wooldridge (2009), which includes an extensive discussion of RED methods and propensity score matching, among other useful methods. For a guide to program evaluation and implementation, see Duflo, Glennerster, and Kremer (2007), which includes a discussion of RED methods.

For convenience of discussion and for some types of analysis, treatments can be divided into two broad categories: event-based and non-event-based. This distinction is useful for determining the appropriate analytical methods. A pilot, however, may deploy any number of treatments that are variations on a basic rate or feedback structure or that involve combinations of characteristics of both event-based and non-event-based treatments.



For event-based treatments, customers are offered inducements to reduce their load during only a few hours of a few days of the year—referred to as events. Those inducements, which are treatments in the experimental vernacular, can involve event price incentives (that is, the potential to reduce the bill) or penalties, as are common in direct load control rate designs to enforce load reduction agreements. The effects of the treatment during the events and perhaps even in the hours immediately surrounding them are expected to be much larger than at any other time. The analysis for estimating the effects of event-based treatments often focuses on the impacts during these event periods and ignores spill-over effects—load changes in other periods of event days. This does not have to be the case, however; the analysis can and should be constructed to also test whether load at other times of the day is affected.

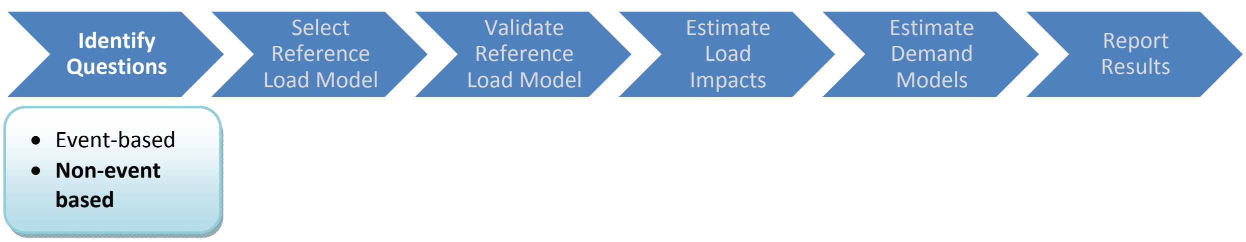
Event-based treatments are typically constructed as overlay options on a basic rate for firm electricity service, which results in temporary and occasional adjustments to the underlying rate structure.[6] For residential customers, there are three main types of overlay options: critical peak pricing (CPP), critical peak rebate (CPR), and utility load control programs (DOE 2010c; EPRI 2011).[7] These are equally applicable to commercial and industrial customers.

Based on short-term system emergency conditions, high costs, or both, a critical peak pricing (CPP) overlay allows the utility to call a limited number of events during pre-specified time periods, such as 1:00–5:00 p.m. on summer weekdays. Under most CPP programs, critical events are determined by system conditions: highly elevated supply costs to avert setting a higher system peak demand, or the possibility that system reliability will be compromised. The CPP event, whose price incentive has been pre-established, is announced to customers in advance; notice may be delivered anywhere from a few hours to the day before the event. Customers pay the higher CPP price for all electricity usage during the critical event but typically pay a slightly lower rate than other customers on a standard rate all other hours of the year.

With a critical peak rebate (CPR), also called a peak time rebate (PTR), pricing overlay is conceptually similar to a CPP except that under a CPR, the participant during an event is paid an

[6] A taxonomy of electric pricing structures can be found in EPRI report 1021962. Overlays, sometime referred to as *riders*, modify some aspect(s) of a basic rate structure, which usually applies only occasionally and under unusual conditions.
[7] Two terms are used to describe the same rate structure: peak time rebate (PTR) and critical peak rebate (CPR).

event credit ($/kWh) for energy reductions measured relative to a customer baseline load (CBL). There is no higher rate or penalty associated with consumption during the critical event itself. A direct load control (DLC) overlay allows the utility to directly control a residential customer's individual appliances, such as water heaters, air conditioners, or pool pumps. In some cases, the load is controlled by the customer, which allows the customer to choose when and by how much to reduce load during events. When the device control is solely under the authority of the utility, there is often an option for the customer to override the control. Although that may seem to be a subtle distinction, it is an important one. There may be a very different event response depending on the nature of human behavior, especially based on the size of the penalty for overriding a control instruction.



Non-event-based treatments provide motivations to routinely change usage behavior on an ongoing, daily basis through treatments such as time-based rates, information and feedback about usage behavior, or control devices that facilitate easier changes in usage behavior. In contrast to event-based treatments, a non-event treatment is always in effect after its onset rather than only during certain critical events. For these types of treatments, it might be reasonable to hypothesize that, in addition to leading to a reduction during certain targeted times of the day, the treatment might also lead to an overall reduction in energy use and a shifting of usage from targeted to non-targeted times of day. The analysis of both of these effects is discussed later in this report.

There are several categories of non-event treatments, including time-of-use (TOU) rates,[8] in-home displays (IHDs), programmable communicating thermostats (PCTs), feedback mechanisms, educational efforts, and combinations of these. For these types of treatments, it may be reasonable to hypothesize that, in addition to leading to a reduction during certain targeted times of the day, the treatment might also lead to an overall reduction in energy use and/or a shift in usage from targeted to non-targeted times of day. For example, a PCT might result in lower settings many times of the day and year that result in reduced electricity usage.

Other time-based rates, such as real-time pricing (RTP) [9] or variable peak pricing (VPP) [10], embody characteristics of both event-based and non-event-based treatments. These rates embody

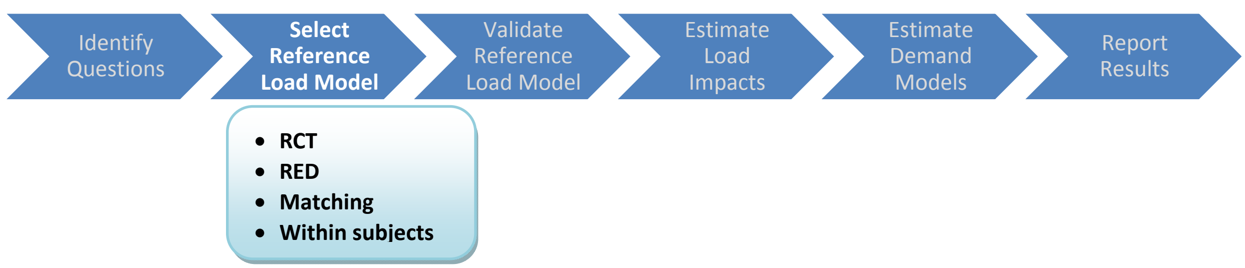
[8] A time-of-use (TOU) rate typically has two or more daily price periods. In the case of two periods, there is a lower commodity price non-peak period and a higher commodity price peak period. In some instances, there can also be what is commonly known as a *shoulder period*. This shoulder period typically is defined as the hours adjacent to and on either side of the peak period, and the price in this shoulder is generally between the non-peak period price and the peak period price. Because prices on peak and on the shoulder period are higher than those during the non-peak period, the TOU rate provides incentives for customers to shift usage away from the peak to the shoulder and off-peak periods and from the shoulder to the off-peak period.

[9] A real-time pricing (RTP) rate provides customers with incentives to shift load from higher priced periods to lower priced periods. Customers face prices set for each hour of the day, and the prices are usually scaled to follow the cost of electricity supply or, in deregulated markets, hourly prices are generally scaled to wholesale energy market prices (for example, PJM Interconnection (PJM) day-ahead or real-time energy market prices). For residential

characteristics of both event-based and non-event-based treatments because prices vary according to prevailing system supply and reliability conditions. At times, prices reach levels close to or commensurate with event prices; at other times, prices are measurably above the average price. Both of these situations offer inducements to alter usage. Furthermore, many other treatments that have recently been used in research efforts likewise embody characteristics of both event-based and non-event-based treatments—as is the case when dynamic rates are combined with overlays.

For example, some of the treatments involving CPP or CPR overlays have flat rates as the background rate, but others have either TOU or RTP rates as background rates. As with non-event-based treatments, these "hybrid" treatments are always in effect, but—as with event-based treatments—they present customers with particularly high prices at times that are not known far in advance. For these types of hybrid treatments, and depending on the utility's objectives, it may be appropriate to estimate event-related load impacts as well as the load shifting and energy conservation impacts that result when other inducements are built into the rate structure.

2.3 Experimental Design and Selection of Reference Load



The quality and usefulness of the estimated effects of electricity rate treatments or other interventions depend significantly on their internal and external validity. If a study is internally valid, it can reasonably be stated that the estimated impacts were caused by the intervention being evaluated. If a study is externally valid, the findings can confidently be extrapolated to a larger population of interest.

The choice of experimental design is the most important factor for establishing internal validity. This choice must be made before the study begins so that customers can be recruited in a way that is appropriate for the particular design. To obtain internally valid estimates of how any treatment affects household-level electricity usage, one needs to compare measured consumption after the intervention with an unbiased estimate of the household-level electricity consumption behaviors that would have been observed in the absence of the intervention. The unbiased estimate of load in the absence of the treatment is often referred to as the reference load, the

customers, these prices are typically supplied to the customers with a day-ahead notice, but they could also be provided with as little as an hour's notice.

[10] Variable peak pricing (VPP) is a hybrid between a TOU rate and an RTP rate. For the VPP rate, the peak period is defined in advance, but the price established for the on-peak period varies daily based on system or market conditions. The same low off-peak price is charged for usage during each hour of the off-peak period.

counterfactual load, or the control group. Internal validity therefore critically depends on the study's experimental design, which determines how the counterfactual is established.

Throughout this report, four basic categories of experimental design (with corresponding reference load methods) are considered: randomized controlled trials (RCTs), randomized encouragement designs (REDs), matched control groups, and within-subject designs.[11] Before continuing, we emphasize that RCTs and REDs produce unbiased estimates of treatment effects that are based on a minimal set of assumptions by virtue of how the reference loads are established. Matched control group and within-subject methods, on the other hand, require much stronger assumptions about the nature of customers in the control and treatment groups. If these stronger assumptions are violated, these methods may produce biased results.[12] The extent of the biases in the estimates of load impacts is not knowable a priori or even after the fact, so there is no way to judge the extent to which the results are misleading. Therefore, during the initial stages in any pilot study's design, we cannot recommend the adoption of either the matched control group or the within-subject design.

Because the purpose of this report, however, is to outline a general set of research protocols to assess the impacts of CBS experiments and pilots, we recognize that there are still circumstances under which these latter two research designs will come into play. Despite efforts to encourage careful consideration and selection of an appropriate research design during the planning stages of a new pilot study, there will continue to be instances in which an appropriate and rigorous design was not adopted. Equally important—and despite the best efforts to identify an appropriate control group—there may be problems in a study's implementation that preclude the control group's use in defining a reference load even though the experimental design was an appropriate one.

A similar situation may arise, for example, if smart meter interval data are available only for customers exposed to the treatment. Under any of these circumstances, a reference load must be constructed another way—perhaps using matching or within-sample methods. In what follows, reasons for these circumstances are discussed in greater detail. We do so because, by not doing so, the only recourse would be to abandon any formal (and structural) evaluation of the data and results—in other words, throwing away the data.

For these reasons, throughout this report a hierarchy of analytical methods to be used to estimate the effects from treatment interventions is discussed. The extent to which the various methods

[11] There is a fifth type of experimental design method called *regression discontinuity*. This method is not described in detail in this report because at this time it is not commonly used in energy evaluations and requires a more complex analysis method. This method works if the eligibility requirement for households to participate in a program is based on a cutoff value for a characteristic that varies within the population but that cutoff value is not endogenously determined. For example, households at or above a cutoff energy consumption value of 900 kWh per month might be eligible to participate in a behavior-based efficiency program, while those with energy consumption below 900 kWh are ineligible. For examples of program evaluations using regression discontinuity, see Ito (2012) and Davis and Boomhower (2013). Also see Imbens, G. W. and T. Lemieux, 2008: "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics* 142 (2): 615–635; Ito and Koichiro, "Does Conservation Targeting Work? Evidence from a Statewide Electricity Rebate Program in California," 2012.

[12] Regression discontinuity designs typically yield results that are less biased than those from matched control group and within-subject designs. See Imbens and Lemieux (2008).

can be used depends, of course, on the initial experimental design, the success in implementation, and the available data. Much of the focus is on the circumstances under which it is appropriate and possible to define reference loads based on RCTs or REDs and the appropriate interpretation of methods and results. When circumstances do not allow for the application of these methods, we suggest methods lower in the hierarchy and discuss how the results must be interpreted differently. This thorough discussion of the analytical methods and their application may also facilitate comparisons with the results from previous studies.

In this section, the discussion is focused primarily on RCTs and REDs. For both event-based and non-event-based treatments, Section 3 describes the detailed mechanics for estimating load impacts based on RCTs and REDs as well as for other methods, such as those that rely on matching and within-subject methods. Section 4 discusses structural economic demand models needed to estimate measures of price response—a critical component in an overall program assessment.

2.3.1 Randomized Controlled Trials

In the program evaluation literature, randomized controlled trials (RCTs) are widely viewed as the "gold standard" of evaluation methods. Their application in CBS experiments is discussed in EPRI report 1020855. The process begins by selecting a random sample of customers from the population of interest (that is, the sample frame). The second step is to randomly assign customers to either the treatment or the control group. To ensure both internal and external validity of the results, it is necessary that this assignment be through conscription. That is, customers in the sample frame have no choice as to whether they are assigned to the treatment or to the control group. Those assigned to the treatment group will be subject to the treatment, and those assigned to the control group will have no access to the treatment.[13] The next step is to administer the intervention of interest to the treatment group. By contrast, the control group receives no intervention. This control group's load becomes the reference load; it provides a measure of the load that would have been obtained for those in the treatment group(s) had they not been exposed to the treatment.

Because this experimental approach eliminates the effects of selection bias and other confounding factors, reference loads constructed in this way facilitate direct comparisons of differences in outcomes across treatment and control groups—the latter serving as the counterfactual—and the estimated treatment effects are unbiased. To maintain the integrity of the design, it is important that no customer choices affect treatment or control group assignments, so that the only difference between the two groups is the result of random assignment.
The main effect of interest for these RCTs is the program impact, or the effect of the treatment program on customers. This is called the intention-to-treat (ITT) effect. Estimates of the ITT effect are derived through simple comparisons of the difference in the outcomes across the treatment and control groups. The mechanics of this analysis for both event-based and non-event-based treatments are described in Section 3 under the RCT subsections.

[13] This is the approach adopted in the large pilot for the Customer Applications Program (CAP) conducted by ComEd in 2011 (EPRI 1023644). Customers randomly selected for the treatment groups were automatically enrolled in the program unless they took specific action to opt-out.

Because this “pure” RCT selects customers for both the control group and the treatment group randomly and no customer is allowed to drop or opt-out of the assignment, there is no volunteer or selection bias—and the internal and external validity of the results is preserved. There are, however, other experimental designs characterized as RCTs, but they differ in some important respects.

As before, the process in these designs begins by selecting a random sample from the population of interest (the sample frame). The second step, however, is to solicit or encourage customers to sign up for the study (usually through an opt-in or an opt-out process); see Figure 1. The third step is to randomly assign customers who already signed up for the study (that is, the recruited customers) to either treatment or control groups. This is done through conscription: once customers sign up for the study, they are assigned randomly and do not have a choice to be in either the treatment or control group. The expectation is that those so assigned will be subject to the treatment.

However, there are several other ways this can be accomplished that do not require the treatment to be mandated. The method chosen affects what analytical methods are used to test if measured changes resulting from treatments are statistically significant.

Figure 2 illustrates the way in which the chosen design points to the appropriate analytical methods to employ.

Figure 1: Recruitment of subjects to CBS experiments: The RED perspective

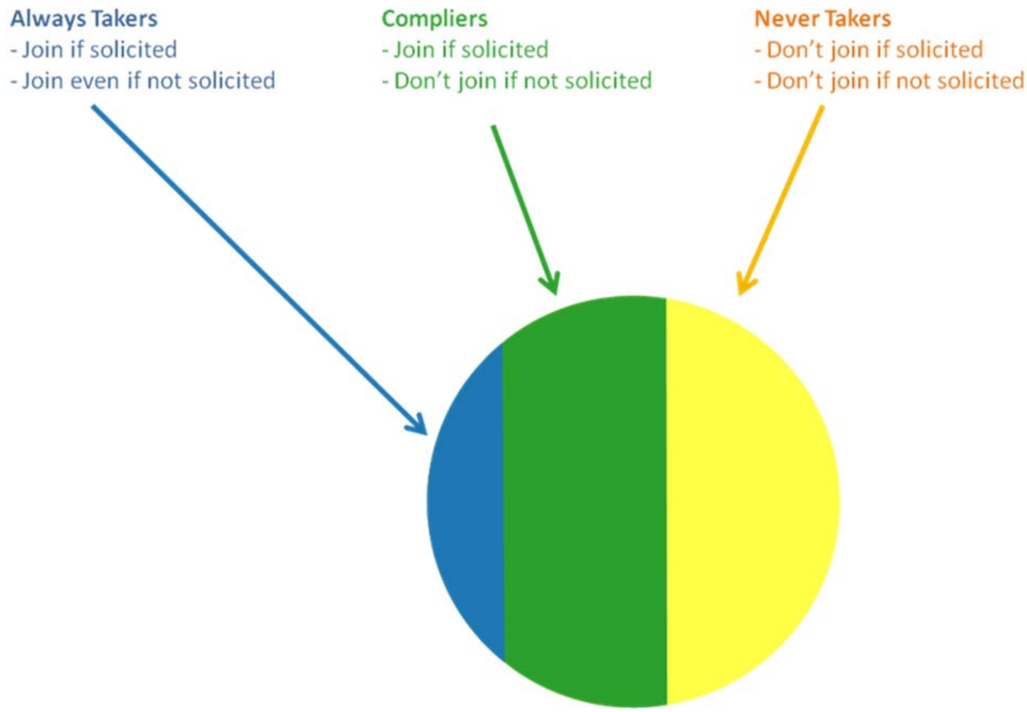
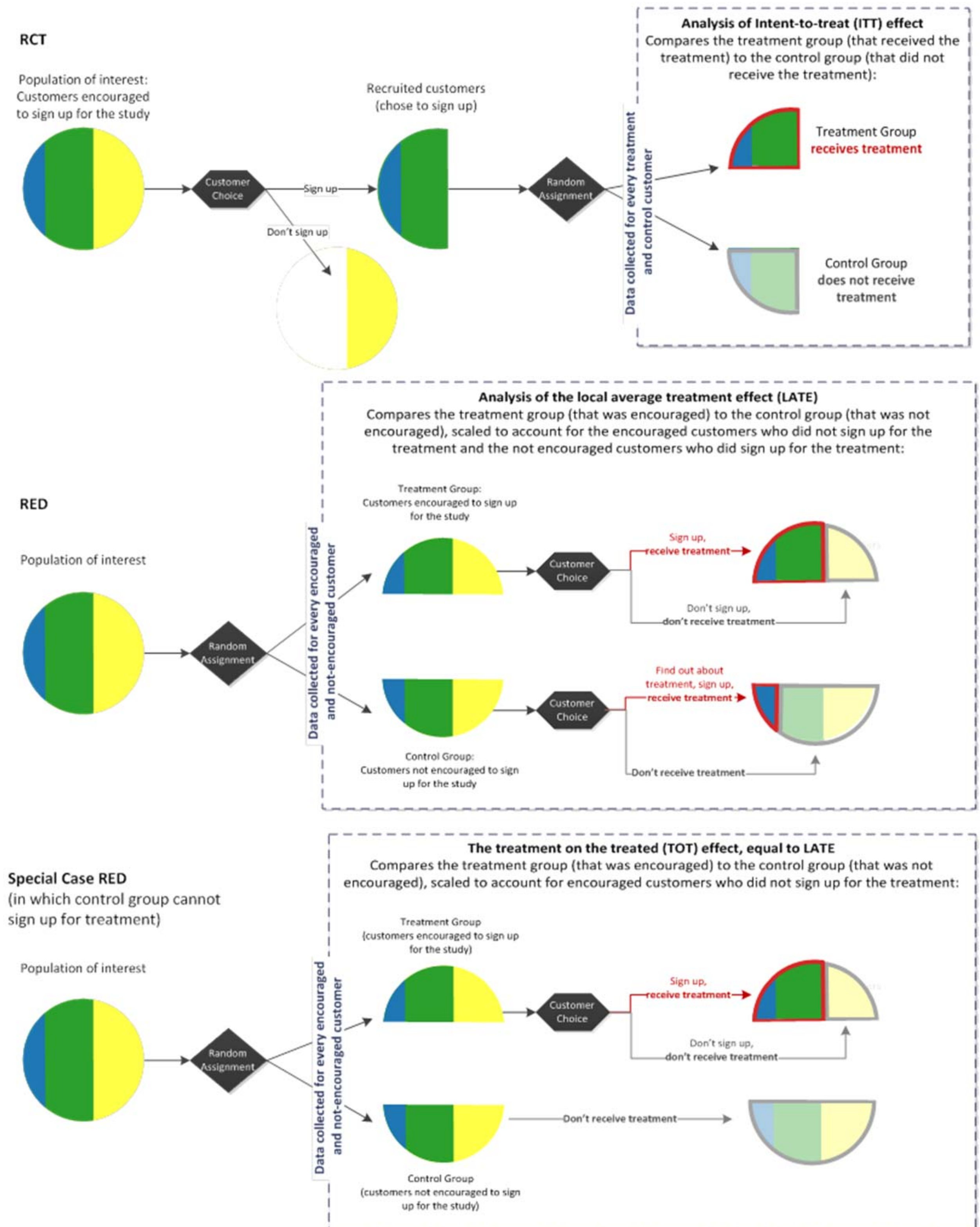
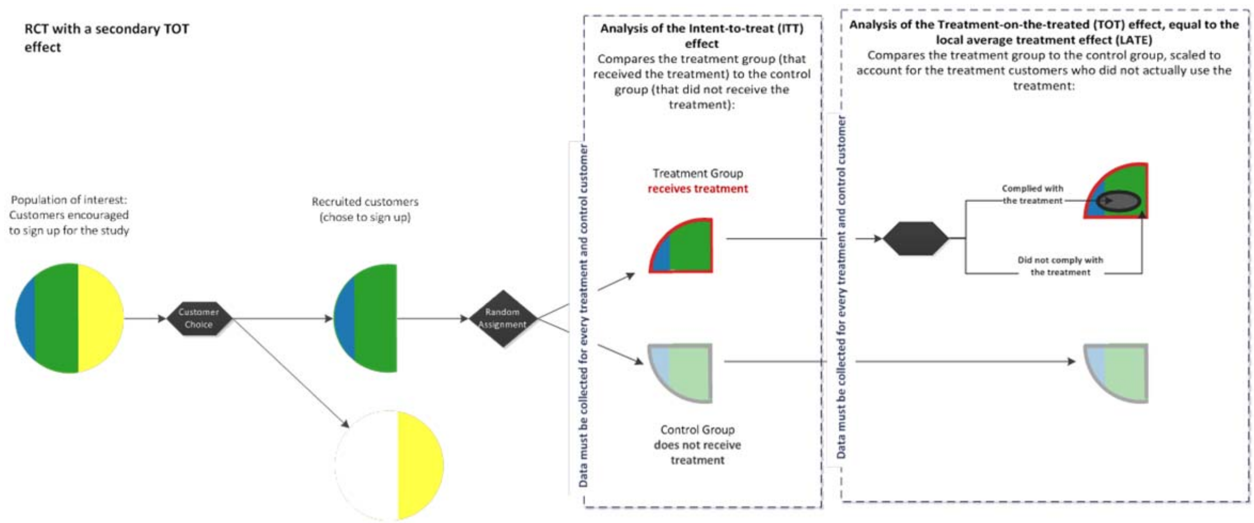


Figure 2: A comparison of experimental designs and their implications for analyzing the results





The fourth step is to administer the intervention of interest to the treatment group. Because, by contrast, the control group receives no intervention, it provides a measure of what would have happened to those in the treatment group(s) had they not been exposed to the treatment.

Because this experimental approach eliminates the effects of selection bias and other confounding factors for customers who have already signed up for the study, reference loads constructed in this way facilitate direct comparisons of differences in outcomes across treatment and control groups—the latter serving as the counterfactual—and result in unbiased estimates of the treatment effects. To maintain the integrity of the design, it is important that no customer choices affect treatment or control group assignments, so that the only difference between the two groups is the result of random assignment. This process guarantees the internal validity of the results for customers who signed up for the study. However, because the random assignment to either control or treatment groups includes only those customers recruited to the program, the results have no external validity.

Recognizing this limitation, this experimental design in other respects resembles that of an RCT. The main effect of interest for these RCTs is the program impact, or the effect of the treatment program on customers who signed up for the program (for example, the effect of an IHD program on customers who signed up for the program by opting in or by not opting out). Again, this is called the intention-to-treat (ITT) effect.

This ITT effect is an unbiased measure of the treatment effect only if one can enforce mandatory compliance with the treatment. In many cases, however, at least a few customers in the treatment group may not actually comply with the treatment (for example, in an IHD study, customers assigned to the treatment group may not actually install and plug in the IHD; in a CPP study, customers may refuse to actually sign up for the rate after they have already been assigned to the treatment group). In this event, the ITT effect is an unbiased measure of being assigned to the treatment but not necessarily complying with it. For this ITT effect to be unbiased, it is necessary to use data from every customer originally assigned to either a control or treatment group, regardless of whether they later dropped out or did not comply.

In addition, there may be a secondary effect of interest: the effect of the treatment-on-the-treated (TOT), or the effect of actually complying with the treatment (for example, the effect of an IHD on those who actually installed and plugged it in or the effect of a CPP rate on those who were actually exposed to the rate).

The ITT effect is estimated simply by comparing the difference in the outcomes across the treatment and control groups; this analysis is described in Section 3 under the RCT subsections. To estimate the effect of the TOT, we can proceed in two steps. The first step is to estimate the ITT effect using all of the customers in the treatment and control groups, regardless of whether they complied. This is followed by a scaling exercise to focus only on the subset of customers that was treated (see Figure 2-2). Because this process for estimating a TOT effect is actually a special case of the estimation used for a RED, it is described in detail in the RED subsection of Section 3. For the more general RED experimental design, this TOT effect is called a local average treatment effect, or LATE. This is described in more detail next.

2.3.2 Randomized Encouragement Design

Despite the considerable advantages offered by RCTs, it may not be feasible to randomize recruited customers who already signed up for the study and place some of them into a control group where the treatment is denied or delayed. In this case, a randomized encouragement design (RED) can be a good option. Essentially, an RED moves the randomization process closer to the beginning of the customer enrollment process so that customers are randomized into a control or treatment group before they are solicited or encouraged to sign up for a study. This means that after randomization, customers still make a choice that impacts their disposition in the study. However, as discussed, this choice does not affect their group assignment for analysis.

Under a RED design, a population of customers is still subject to random assignment to either a treatment or control group, but the treatment consists of an encouragement to take up the rate or technology treatment through a subscription process that allows customers to decline participation (see Figure 1). For example, the encouragement might consist of a marketing program implemented to a randomly chosen subset of the eligible population to recruit customers to a time-based retail rate. The control group consists of randomly chosen customers who are not encouraged to take up the treatment. In most cases, the control group customers do not know that they are involved the study (therefore, attrition occurs for reasons external to the study). If control group customers learn of the opportunity afforded by the treatment on their own, they may be allowed to take up the treatment. One special case of a RED design is if the control customers are not allowed to take up the treatment even if they learn of it on their own. This special case may in fact be the most typical case for energy pilots.

By directing the encouragement at only a subset of the population, more customers in the treatment group to whom the encouragement is directed will, in all likelihood, sign up for the treatment than will do so in the control group (especially for the case in which control group customers are not allowed to take up the treatment). Therefore, the application of an encouragement induces a correlation between being in the treatment group (for example, being asked to join the rate) and signing up for the treatment (for example, being on the rate). This correlation can be used to produce an unbiased estimate of the effect of the treatment on the customers who subscribed to it because of the encouragement, termed a local average treatment effect (LATE). For the special sub-case in which control customers are not allowed to take up the treatment, this unbiased LATE estimate will be the effect of the TOT and will give the same result an RCT would have given.[14] The analysis of the LATE estimate can be accomplished in two steps, the first of which is the same as the analysis of the ITT estimate described in Section 3 under the RCT subsections, and the second of which is described in Section 3 under the RED subsections.

2.3.2.1 A Further Discussion of the LATE and TOT Effects

Although the mechanics of estimating the ITT and the LATE and TOT effects for both event-based and non-event-based electricity rate treatments are discussed in Section 3, it is critical at this point to understand the assumptions that underpin the validity of these estimators, the frameworks that can motivate the estimators, and some practical issues about the extent to which the data needed to estimate these effects can be made available during the conduct of pilot rate programs.

To help motivate the discussion, outlined below is a simple instrumental variable (IV) approach to the estimation of TOT and LATE effects.[15] This simple model is adapted from Brookhart et al. (2010). The ITT estimator is given in Eq. 1. When the compliance guarantees needed for the ITT estimator to be an unbiased estimator of the load impact are not possible, Brookhart et al. (2010) argue that the ITT estimator can at best provide evidence about the presence of the treatment effect. They go on to provide an instrumental variable estimator of the TOT or LATE effect. They consider the case of a dichotomous instrument and demonstrate the connection to a linear structural equation model in which the TOT or LATE effects can be estimated by two-stage least squares or an equivalent two-step application of ordinary least squares.

In an RCT, the ITT estimator for mean outcomes between rate treatment and control groups is shown in the following equation:

Eq. 1: $\beta_{ITT} = E[Y|Z = 1] - E[Y|Z = 0]$

where: E is the expectations operator, Y is the outcome (for example, load impact), and Z is the instrumental variable; Z = 1 if the customer is assigned to receive the rate treatment and zero otherwise. If one is guaranteed that control group customers have no access to the rate treatment and that all customers in the rate treatment group actually go on the rate, this ITT estimator provides an unbiased estimate of the treatment effect—the load impact.

If such compliance guarantees are not possible, this ITT estimator can at best provide evidence only about the presence of the treatment effect. When no such guarantees are possible, treatment

[14] This is one reason that an RCT is preferable to an RED. The estimated effect for an RCT can always produce estimates of the treatment-on-the-treated, whereas if an RED allows control customers to take up the treatment, the RED can produce an estimate of the treatment only on the customers who took up the rate because of the encouragement, which leaves out the treatment on those who took up the rate for other reasons.
[15] The source for this example is Brookhart et. al. (2010); adapted from the discussion of the model applied to medical interventions.

effects can be modeled in an instrumental variable framework. With a dichotomous instrument, the IV estimator of the TOT or LATE effect is as follows:

Eq. 2: $\beta_{IV} = \frac{E[Y|Z=1]-E[Y|Z=0]}{E[X|Z=1]-E[X|Z=0]}$

where: Y again is the outcome, Z is the instrument, X is the rate treatment, and β_{IV} is a measure of the effect of X on Y. The numerator β_{IV} is the ITT estimator, but the denominator is the difference in treatment rates between levels of the instruments—a measure of compliance. If there is perfect compliance or if we are in the guaranteed conscription case, $E[X|Z=1]-E[X|Z=0]=1$.The instrument predicts the rate treatment exactly, and the IV estimator is equal to the ITT estimator. As noncompliance increases, the denominator of the IV estimator shrinks and the IV estimator itself increases relative to the ITT estimator.

This simple estimator, β_{IV}, is called the *Wald estimator*. It can be motivated in a variety of ways—one is through its connection to a set of linear structural equations. Interpreted in this way, there is a two-equation model. The first equation (X) is one for the rate treatment assignment that is a function of the instrument and perhaps potential confounding factors. The outcome equation (Y) is a function of the treatment and potential covariates.

Eq. 3: $X = \alpha_0 + \alpha_1 Z + \alpha_2 C + \epsilon_1$

Eq. 4: $Y = \beta_0 + \beta_1 Z + \beta_2 C + \epsilon_2$

where C is a vector of possible covariates, and α's and β's are vectors of parameters. The treatment effect is β_2. The ordinary-least squares (OLS) estimate of this effect is unbiased only if the exposure to treatment is uncorrelated with ϵ_2. The magnitude of the bias is directly related to the strength of this correlation.

These two equations can be estimated by two-stage least squares or equivalently in two steps. The equation for X is estimated by OLS. The equation for Y is estimated by OLS after replacing X with its predicted values from the estimated equation for X. If there are no covariates included in these two equations, the resulting two-stage least squares estimates of β_2 is equal to the Wald estimator.

Table 1 and Table 2 offer side-by-side comparisons of assumptions and the data requirements needed for unbiased TOT and LATE estimators. We recommend that the material in these tables be studied carefully. This side-by-side comparison is particularly convenient because of the similarity between a study designed as the special case of an RED that does not allow control customers to take up the treatment and a study designed as an RCT with a secondary TOT effect for treatment customers who actually use the treatment. In the RCT case, the assignment to the treatment group could be considered the encouragement, with customers deciding whether to actually use the treatment. In both cases, the first step of analysis is an estimate of the ITT effect, followed by a LATE analysis (which is equal to the TOT effect analysis in both cases). The major difference may be that a study designed as an RED will have to be planned so that enough customers are enrolled to meet the larger sample size needed to provide estimates of the LATE effect with an acceptable degree of statistical significance. If the study were designed initially as an RCT, the sample may not include enough customers to estimate the TOT effect with any acceptable degree of statistical significance.

There are three important assumptions that must hold for the TOT and LATE estimators to be unbiased. These are listed side by side in Table 1because they are quite similar for both the TOT and the LATE estimators. For both, the independence assumption is usually true by design, and the monotonicity assumption is fairly easy to believe. The exclusion assumption, however, may be a concern in the case of a RED design. That is, there may be some cases in which the encouragement itself, separately from the treatment, actually causes energy savings. The best way to deal with this is to design the encouragement so that it emphasizes the importance of signing up for the program rather than emphasizing how important it is to save energy.

The interpretation of the TOT and LATE estimators within a linear structural model framework (for example, Equations 3 and 4 in Table 1) is convenient. When these three assumptions are valid and there are no covariates included in either of these equations, the treatment effect, β, from Equation 4 is identical to the Wald estimator in Equation 2. When there are heterogeneous treatment effects (for example, load response depends on income, house, and family size), however, this linear model may no longer be a good model for the data. Fortunately, even if the treatment effects are heterogeneous (Brookhart et al., 1996), show that under the “monotonicity” assumption, the Wald estimator in Equation 2 still yields the average treatment effect among the “compliers.” This result has important implications for the internal validity of this measure of the average treatment effect.

Table 1: Assumptions to ensure the validity of TOT and LATE estimators

TOT	LATE
Independence	
Customers are randomly assigned to the control and treatment groups.	Customers are randomly assigned to the control group that was not encouraged or to the treatment group that was encouraged.
Exclusion	
Assignment to the treatment and control groups does not affect the impact of using or not using the treatment (for example, it assumes that customers who receive the IHD but do not install it do not save energy just by virtue of having the IHD in an unopened box in their house).	Assignment to the encouraged and not encouraged groups does not affect the impact of receiving the treatment (for example, it assumes that customers who are marketed to and asked to sign up for a time-based rate do not save energy just by virtue of being marketed to).
Monotonicity	
Customers in the treatment group are not less likely to use the treatment than if they had not been in the treatment group (for example, it assumes that although being in the treatment group may have no effect on whether a customer installs and plugs in their IHD, there is no one who did not install and plug in their IHD because they were in the treatment group).	Customers who are encouraged are not less likely to receive the treatment than if they had not been encouraged (for example, it assumes that although being encouraged to sign up may not increase the likelihood that a customer signs up, there is no one who did not sign up specifically because they were encouraged).

Table 2: Data requirements to estimate the TOT and LATE effects

TOT	LATE
An analysis of a study designed as an RCT must include data on the outcome being studied (for example, energy use) for every customer who was originally randomly assigned to a control or treatment group.	• An analysis of a study designed as an RED must include data on the outcome being studied (for example, hourly energy use) for every customer who was originally randomly assigned to the encouraged or to the non-encouraged group, regardless of whether they actually signed up.
Specifically, if some treatment customers drop out of the study midway through, one cannot simply discard those customers; they must be included in the final analysis as part of the treatment group.	Specifically, one cannot simply discard the customers who were encouraged but did not sign up for the treatment. If these customers are discarded, the treatment and control groups would no longer be comparable. The treatment group would contain only the types of people who would sign up for the study, while the control group would contain both the types of customers who would sign up and the types who would not sign up. The control group would therefore no longer be a good reference load for the treatment group. The only exception to this rule is the sources of drop-outs of the utility billing system that can be reasonably expected to affect the treatment and control groups equally (for example, customers who move away—it is unlikely that a customer would be more likely to move away as a result of treatment or non-treatment).
Likewise, for a TOT analysis, one cannot simply discard the customers in the treatment group who did not actually use the treatment. If these customers are discarded, the treatment and control groups would no longer be comparable. The only exception to this rule is that there are some sources of drop-outs of the utility billing system that can be reasonably expected to affect the treatment and control groups equally (for example, customers who move away—it is unlikely that a customer would be more likely to move away as a result of treatment or non-treatment). Excluding these customers affects the study's external validity but not its internal validity.	Customers in the treatment group who do not sign up for the treatment (even though they are encouraged) and customers in the control group who sign up for the treatment (even though they are not encouraged) add noise to the analysis without providing information about the treatment of interest.
It is important to note that many technology-based treatments will likely include several customers who do not actually use their treatment despite the intent to do so. This may be the result of technology compatibility issues. For example, a typical in-home display or programmable thermostat will have certain limitations that allow it to work only at certain homes. These limitations cannot always be perfectly predicted prior to customer recruitment. Therefore, some customers	Estimates of LATE therefore have greater variance per treated customer than do ITT estimates; therefore REDs require a larger sample size than RCTs.

will be assigned to the treatment group, but they will be unable to have a device installed. In this situation, those customers should be kept in the treatment group to maintain the original comparability between treatment and control groups. The impacts can then be estimated using an ITT analysis followed by an analysis of the TOT.	

Because a major purpose for conducting a large-scale rate pilot is to estimate load impacts, hourly load data by customer is typically among the data collected throughout the study's treatment period. The major challenges to be overcome in preparing these data sets for any empirical estimation are well known and similar to those that would be encountered in any large-scale empirical analysis.

There is yet another essential data requirement for estimating TOT or LATE effects. In both cases, data on the outcome being studied (for example, hourly load data needed to estimate the load impact) must be obtained for every customer who was originally randomly assigned—to a control or treatment group in a TOT analysis and to the encouraged or non-encouraged group in a LATE analysis. These data must be collected regardless of whether the customers actually signed up for the rate treatment.

The discussions in Table 2 underscore reasons these data must be collected for every customer who was randomly assigned as well as the consequences for not having the data. It is particularly critical to collect these data for customers who were either randomly assigned to the treatment or encouraged to do so but chose not to. Otherwise, as explained in Table 2, the control and treatment groups will no longer be comparable.

Despite the compelling reasons for needing these data, these data requirements potentially pose serious challenges in the implementation of any pilot study. If an appropriate meter is installed, there may be little difficulty in collecting load data for customers assigned to the treatment group but who were not able to participate because a critical device (such as an IHD or a controllable thermostat) could not be installed. This is also likely to be the case in an RED for customers in the non-encouraged group who are offered the treatment after they have contacted the utility.

The situation can be quite different in pilots in which the utility must install the meters needed to collect the hourly load data both for customers in the control group and those randomly recruited for the treatment group using an opt-in strategy. Based on experience in past pilot studies, utilities are not inclined to install meters for customers randomly recruited but who did not opt-in to the treatment. It raises the cost with no apparent benefit. In these cases, there seemed to be no reason to be concerned about metering the load for customers who did not opt-in; therefore, the expense of installing the meter could hardly be justified.

On the surface, there might seem to be an easy remedy: make those in charge of pilot implementation aware of the issue, and have them make sure that meters are installed. The larger problem, however, stems from the fact that in past pilots, only 5-15% of those customers contacted actually opt-in. This means that meters must be installed for the remaining 85–90% of those customers contacted initially. Given the low opt-in percentages, it is easy to envision how the expense of meter installation could be prohibitive. If the expense is not an issue, and the meters are installed, however, these customers are still "non-compliers." This 85–90% non-compliance rate could, in turn, affect the validity of estimates of the TOT or LATE effects. This happens because of the effect of these high rates of non-compliance on the performance of the instrumental variable equation (Equation 3 in Table 2-1). That is, the instrument (Z) in this equation may now not be a strong predictor of who agreed to go on the rate treatment, Brookhart et al. (2010, p. 540) point to three problems with weak instruments:

- The finite sample bias of standard IV estimators is inversely proportional to the F-statistic that tests for the significance of the instrument in the first-stage model.
- Small violations of the IV assumptions can lead to tremendous inconsistency in the IV estimator when the IV is weak.
- Weak instruments yield highly variable estimates of treatment effects and limit the power to detect small effects even in very large studies.

2.3.3 When a Control Is Not Part of the Design

In the evaluation of some CBSs, it may be impossible to use either an RCT or a RED method to develop a reference load, for several reasons. For example, smart meter interval data may be available only for customers exposed to the treatment because the experiment commences with the installation of the meters. Or, despite the best intentions, there may have been unforeseen difficulties in selecting the control group. These difficulties may be uncovered only after the fact through a careful examination of data for the control group; these difficulties may lead to the conclusion that it is impossible to define an appropriate reference load based on the control group.

Under these circumstances, there are two main options—both of which rely on making assumptions in addition to those required for either a RCT or a RED. First, it may be possible to construct a matched control group using customers from the broader population. Second, it may be possible to use the load observed for the treatment customers during non-event periods to define a reference load model and thereby produce an estimate of the effect of an event-based treatment.

In a recent EPRI report on the evaluation of energy information feedback pilots, the following conclusion was stated with respect to the use of matched control groups:

> *"Matching methods by themselves are to be used sparingly because they are prone to the introduction of bias that cannot be anticipated or measured. The calculated estimates of differences (or difference of differences) are biased (they cannot be inferred to reflect the real values) and inconsistent (the variance is large and unknown, so we cannot make statements about the confidence interval around the estimate). These constitute a strong cautionary. However compelling the results based on experience, intuition, or other indicators of a treatment effect, the experiment does not provide confirming and incontrovertible evidence that the observed effect is attributable solely to the treatment"*

That conclusion was specific to the use of matching for information feedback, a particular type of non-event-based treatment. The same conclusion applies here, however, with a caveat: in certain cases of event-based treatments with sufficiently rich analysis data, the expected treatment effect is large enough and the range of likely bias small enough that a matching analysis may be useful. A similar point applies to within-subjects analyses, discussed next.

The only situation in which either of these methods is convincing enough to use as a reference load is in the analysis of an event-based program where there are a substantial number of non-event periods with conditions similar to those observed during events (for example, similar temperatures or similar time of the week). Without these conditions, these methods are suggestive at best and should not be used for decision making that requires a high degree of confidence. This discussion of these methods is limited to this situation.

The assumptions underlying each of these options are somewhat similar. At a high level, they are as follows:

1. For within-subjects designs, we must first assume that there is no selection bias in the customers who subscribed to the study.
2. We must assume that we can control for all differences that exist between event periods and non-event periods using only observable variables. For example, it is frequently the case that event days coincide with very hot days and that the set of non-event days contains few or no days as hot as the event days. In this case, we must assume that we can control for the differences in usage that arise between event and non-event days—aside from the effect of the events—using a function of temperature. In other cases, event days may not coincide with the hottest days, but we must always look for ways in which they differ systematically from non-event days.
3. For a matched control group, we must first assume that we can control for all selection bias using observable variables (such as demographic and geographic variables as well as loads observed during non-treatment periods). Second, we must assume that we can also control for all differences that exist between event periods and non-event periods using observable variables and the change in energy use during those periods for the matched control group. For example, on hot days, we must assume that households may change their energy use relative to non-hot days, but the treatment and matched control groups would have changed their usage because it was hot in the same way in the absence of an event after controlling for observables.

For matched (after the fact) control groups, the most promising variables on which to match are loads observed during non-event periods with event-like temperature conditions. For within-subjects methods, a credible reference load model for events will require the observation of loads observed under very similar conditions. In each case, there is an implied assumption that customer usage during these non-event periods is an accurate reflection of what usage would have been during event periods if the customer had not been on the treatment. If event days are not systematically different from non-event days—except in ways that can be controlled for—that may be a reasonable assumption. If non-event day periods during the post-CPP enrollment period are used in the analysis, both matched control groups and within-subjects analyses also require the assumption that the treatment affects behavior only on event days; therefore, non-event periods provide a useful counterfactual.

For matched control groups, regardless of how well the matching procedure appears to work (that is, when the selection criteria are compared, the control and treatment appear identical), the possibility of unobservable differences between groups producing bias in the results cannot be ruled out. This is frequently described as relying on an unconfoundedness assumption, whereby we assume that there are no variables that are unobservable and correlated with both the treatment and usage behavior in the population. This is a strong assumption that typically cannot be tested (unless the study is specifically designed as a RCT and analyzed with a matching method so that the degree of bias can be estimated). The implication of this assumption is that standard judgments about statistical significance of estimated impacts must be qualified: statistical significance is always conditional on the unconfoundedness assumption. In the context of an event-based program, in which non-event period load is used for matching, the unconfoundedness assumption can be expressed as follows: just because the matched customers have loads similar to those of the treatment group during event-like periods does not mean that they will have loads similar to those of the treatment group during actual event periods. This risk is unavoidable with this method, and the consequences are unknowable—therefore the admonishment to avoid this recourse by adopting a RCT design.

The matched control group method assumes that the way in which the treatment group's energy use would have changed in the absence of an event is fully represented by the control group and observable variables. Within-subjects methods require the additional assumption that the way in which the treatment group's energy use would have changed in the absence of an event is fully represented by observable variables only. For example, it assumes that a proper functional form can be developed for expressing load as a function of temperature, day of the week, and other observables in order to predict reference load (for example, a fixed effects model, discussed next, that includes many other factors that influence load and therefore net out the event effect.) Therefore, within-subjects methods have additional specification error risk. This risk is likely small if there are many non-event days with event-like conditions. The risk is unknowable otherwise and may result in the drawing of erroneous conclusions that have adverse consequences. The fact that a pilot is being undertaken suggests that there are important uncertainties that need to be resolved. Additional risk would not contribute to the goal.

2.3.4 Summary

In this section, we have laid out the required assumptions and measurements necessary to implement a valid RCT or RED. The best method is an RCT. However, this requires assigning customers to treatments, which may not be feasible or practical, thereby diluting the statistical basis for the findings (EPRI report 1023644, 2012). A common approach is to invite customers to participate, focusing the analysis on how volunteers respond to treatment. The results apply only to the population of customer that would volunteer. However, if the program being tested is intended to be employed using self-selection, this approach comports with the project goal.

The common practice is to focus the measuring of electricity usage on those who volunteered to participate. In the last subsection, we raised an issue that should be of concern for utilities that may have undertaken pilot studies without metering those customers who declined the invitation to participate in a treatment or who turned out to be non-compliers among the opt-in treatment groups. Under some circumstances, the control may be corrupted, despite its having been drawn randomly from the population of potential volunteers (the sample frame).

The key to a credible conclusion about the treatment effect is establishing a control group under an RCT or a RED. If a randomized control is not drawn at the onset or data are not correctly collected, as discussed elsewhere in this report, we must in this case resort to the non-experimental methods of matching and/or within-subjects methods to construct a reference load. Such an analysis based on either of these methods may never be free of the possibility of bias. Despite this potential difficulty, the only recourse (that is, to not conducting the analysis based on one of these two methods) would be to abandon any formal (and structural) evaluation of the data and results—throwing away the data.

2.4 Assessment and Diagnostics of the Research Design

Every research design has potential flaws. The experimental designs recommended in this report—RCTs and REDs—typically have the best chance of producing accurate impact estimates. However, even they can lead to inaccurate results in at least two ways. First, they can be incorrectly implemented. Second, although randomization ensures that treatment and control groups are identical in expectation and that over many experiments they are identical on average, in any implementation of an experiment the treatment and control groups will be less than perfectly balanced across observable variables. This is not an indication that the experiment or the impact estimates are invalid. However, these differences can affect the outcome of the experiment and are therefore important to understand in interpreting the results of a given implementation of an experiment.

In the evaluation literature, this issue is sometimes referred to as the amount of overlap in covariate distributions between treatment and control groups. In the ideal experiment, the overlap between the distributions of covariates in treatment and control groups would be complete. It is simple to show empirically, however, that some randomly drawn groups will be more like one another than others. Because the control group is meant to provide a view of the behavior of the treatment group in the absence of the treatment, a study is more likely to produce impact estimates reflecting only the effect of the treatment when the characteristics of the treatment and control groups are more similar prior to treatment.

Imai, King, and Stuart (2008) state the issue this way:

> "Randomly assigning the values of the treatment variable…which is normally considered the sine qua non of experimental research, reduces the components of estimation error

arising from observed and unobserved variables on average, but not exactly in sample… [16]

Similarly, Altman (1985) says:

> *"While there is a probabilistic argument that on average random groups will have the same characteristics, it is quite likely in practice in a given trial that the subjects with some particular characteristics will not be split equally between groups.*
>
> *Imbalance between baseline groups is clearly undesirable, as the essence of a controlled trial is to compare groups that differ only with respect to their treatment. "*

As a simple example, suppose we use an RCT to estimate the impact of a CPP rate. Suppose that through random selection, the treatment group has 10% greater prevalence of air conditioning (AC) in the home. This is still a fully valid experiment, and its results are properly considered among the results of a set of other CPP experiments. However, in interpreting the results of this experiment in isolation, it is useful to know about the difference between the treatment and control groups. In this case, it would probably suggest that load impacts would be somewhat underestimated because the control group would have a lower load during the CPP period than the treatment group's expected load would have been in the absence of the event.

This bias can be at least partially addressed through the inclusion of pre-treatment load in the analysis and/or through the use of control variables, such as the presence of AC. By estimating the regression models using a "fixed-effects" estimator to capture customer-specific effects, we may be able to account for some unexplained variation in the model. However, these methods can correct for differences only between groups that do not change over time, which, for example, would not perfectly account for the effect of one group having greater AC load. More complicated modeling to account for non-comparability, such as specifying load as a function of recent temperatures and daily and seasonal patterns, may be limited because of lack of data for modeling; it may also lead to misspecification. Such modeling can certainly improve estimates, but the degree of accuracy becomes much more difficult to assess. Therefore, estimates of load impacts are more reliable if the control group is quite similar to the treatment group.

This section describes two simple ways to illustrate comparability between groups. Unfortunately, there is no simple answer to the question of how similar the treatment and control groups must be in order to trust a study's results. Randomly drawn groups are an important foundation, but beyond that, groups that are more alike are preferred. There is no "yes/no" mathematical test to determine whether two groups are sufficiently similar. There are ways to examine the data and tests that can be performed that can bring to light ways in which the groups differ, but whether those constitute reasons to rethink the study's results is at least partially a subjective determination.

[16] "Misunderstandings between Experimentalists and Observationalists about Causal Inference," by Imai, King, and Stuart, published in the *Journal of the Royal Statistical Society, Series A*, 2008.

Therefore, we view the recommendations in this section as good faith gestures of transparency rather than as strict tests of validity. They allow outsiders to assess the possibility that the experiment was not implemented correctly—by checking, for example, whether gross imbalances arise between large treatment and control groups. They also allow outsiders to personally judge whether the random variation between the groups is of concern in interpreting the results of the experiment, even in cases in which the experiment was perfectly implemented.

For analysis types not based on RCTs and REDs, similar assessments of the research design can be performed. With a matched control group method, the treatment group can be compared to the matched control group. For a within-subjects method, the analogous comparison is between actual electric load and model predicted load on non-event days (as described next). In the case of matched control groups or within-subjects methods (for which only one of the following exercises applies), the suggestions can be seen as a "full-disclosure" exercise in which evaluators give reviewers information about how well their modeling eliminated observable bias from the analysis (or how well the modeling produced similar covariate distributions across groups), recognizing the limitations of the methods previously discussed. For example, it would be much more valuable to an outside reviewer to know that a matched control group had pre-treatment differences from the control group of about 50 kWh per month in monthly usage than to simply know that a matched group was created. The methods suggested are not meant to be the "last word" in reference load assessment or experimental best practice, and more statistically sophisticated evaluators may be able to provide better metrics of reference load validity. However, current experience with evaluations suggests that the alternative to these suggestions is typically no assessment rather than more sophisticated assessment. Moreover, the following suggestions have become standard practice to report in the "Data Overview" section of academic reports on field experiments. The main addition that we suggest—graphing pre-treatment loads (or non-event period loads, in some cases)—is a variation on the same exercise but adapted to the field of electric load studies.

2.4.1 Illustrating Comparability Between Groups

We recommend using two main types of data to assess the comparability between groups: demographic data and usage data, preferably at the hourly level. We also recommend tracking and reporting the number of customers in each group who close their accounts during the experiment.

In a recent RCT-based evaluation of an AC load control program, for example, the authors began this validation process by creating a table that shows the fractions of the treatment and control groups located within each of seven local capacity areas—an important geographical characteristic. The table also shows average monthly usage for the summer months for both groups and the fractions of customers in each group who are on PG&E's CARE tariff, an underlying rate for low-income customers (see Table 3). The table contains t-statistics and p-values for the differences between the groups and shows that the RCT sampling worked quite well. Even in cases in which there are statistically significant differences between groups, the differences are quite small on an absolute basis. This is perhaps a classic example of the criticism of p-values: they are mainly a test of whether the sample size is large enough. In this case, we conclude that the sample was well randomized, but the sample sizes are so large (roughly 15,000 treatment and 150,000 control) that small differences sometimes produce significant p-values. This is a valid criticism of reporting p-values in this case; however, in cases of smaller samples and larger deviations in mean values, the p-value is a useful heuristic to help determine how likely it is that the difference is systematic or of concern.

Table 3: Validation of treatment and control groups

Variable	Control	Treatment	T-stat	P-value
Greater Bay Area	0.34	0.34	1.5	93%
Greater Fresno	0.19	0.19	-2.2	1%
Kern	0.04	0.03	6.1	100%
Northern Coast	0.06	0.06	-3.1	0%
Other	0.17	0.17	-2.3	1%
Sierra	0.11	0.11	1.1	87%
Stockton	0.09	0.09	2.5	99%
% of CARE customer	0.29	.29	-1.6	5%
Avg. Summer Monthly (kWh)	743	741	0.3	63%
Avg. Usage from 2-3 PM	1.47	1.45	1.3	91%
Avg. Usage from 3-4 PM	1.76	1.73	1.4	92%
Avg. Usage from 4-5 PM	2.00	1.98	1.0	85%
Avg. Usage from 5-6 PM	2.15	2.15	0.2	59%
Avg. Usage from 6-7 PM	2.10	2.11	-0.5	32%
Avg. Hourly Usage (kWH)	1.15	1.15	0.1	56%

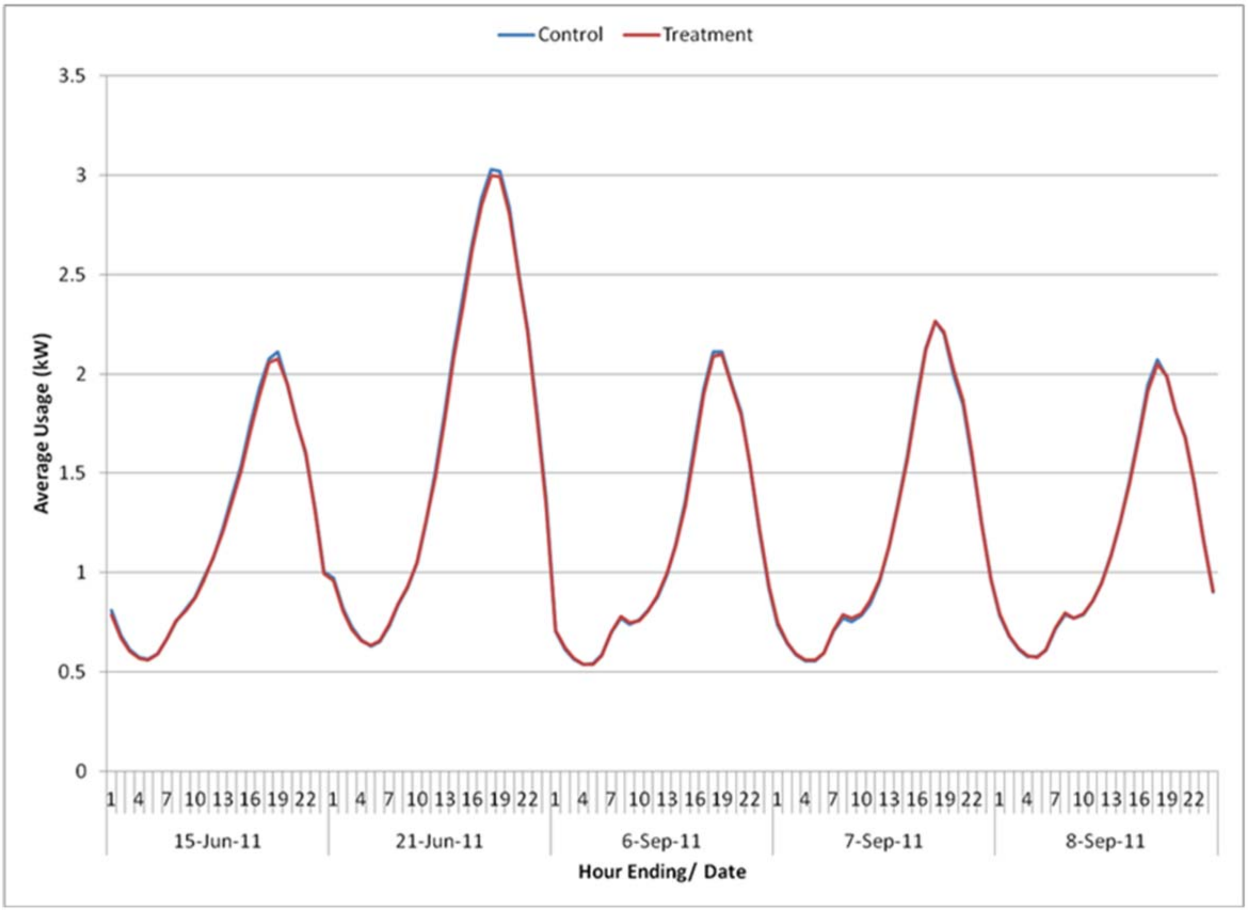
It is, however, more important to demonstrate that the treatment and control groups have similar usage patterns than it is to demonstrate that the groups are comparable across geographic and demographic characteristics. The fundamental assumption underlying the reference load model for each of these analyses is that the control or non-encouraged control group load is an accurate estimate of what the load in the treatment or encouraged group would have been had it not been exposed to treatment. To show that this is likely a realistic assumption, we must demonstrate that loads between the groups are very similar prior to the treatment, under conditions of interest—either event-like conditions for event-based treatments or peak periods for non-event treatments. We recommend showing graphs of hourly average usage in each group for the same set of pre-treatment days that will be used to control for pre-treatment differences in the main analysis (see Section 3 for more discussion of selecting pre-treatment data). It is likely that several of the important graphs may already have been constructed during the initial descriptive analysis of the data.

If no pre-treatment data are available, and if the background rate for a CPP treatment (for example) is the default residential rate, it may be possible to assume that treatment and control customers would exhibit similar behavior on non-event days. The use of this assumption is not recommended unless no pre-treatment data are available. For non-event-based programs, such an assumption is not an option—and little can be done unless pre-treatment data are available. One can hope that in either case the graphs show that the load is generally quite similar between the groups and that it deviates noticeably between groups at some times of the day. Simply seeing the scale and frequency of the deviations can be useful in understanding the degree to which treatment and control groups match, which can be important for interpreting the impact estimates.

For within-subjects analyses, a similar exercise can be performed in which the reference load model is used to predict load during the most event-like non-event periods in the data. This is best done by withholding the event-like conditions from the model for fitting and then predicting the reference load for the withheld periods. To provide the model with some event-like conditions for fitting, we recommend a procedure in which several event-like days are chosen for validation; the model is then fit several times, each time withholding one of the event-like non-event days and predicting for that day. We refer to this as out-of-sample testing. A graph such as the one we suggest to demonstrate treatment and control comparability can then be used to demonstrate the accuracy of the within-subjects reference load model. Although there may be other useful ways to perform this demonstration, this procedure is recommended because it is fairly simple and reduces the likelihood of over-fitting the model (by making it more difficult to do so), which is a frequent problem in within-subjects models.

An example of such a graph is shown in Figure 3 (this is also taken from the RCT-based analysis that produced Table 3). The graph shows that on several hot non-event days, the treatment and control groups have virtually identical loads. This provides strong evidence that on event days, these groups would also have similar loads were it not for the intervention of load control.

Figure 3: Validation of reference load for air conditioning load control



For the case of a non-event-based treatment, it might be convenient to show the average loads over different types of days rather than every day individually. Unlike with the event-based case, for non-event-based treatments we are interested in group comparability under many different conditions. A good way to organize the figures might be according to categories of high temperature on weekdays and weekends. For example, one could show average hourly usage in each group on weekdays with high temperatures in the 70s, 80s, 90s, and above 100—each as its own daily average—and similarly, but separately, for weekends.

This examination of the load data and other observables is of great value in an informal sense because it gives the analyst a much better idea of how comparable the groups are and can indicate problems in implementation. However, the data can also be used to test more formally whether important differences exist between the treatment and control groups. We propose one such test. Note, however, that the assumption of balanced identical underlying distributions of characteristics between treatment and control customers implies dozens of possible statistical tests that could be performed.

In the suggested test, a discrete choice regression (either a probit, logit, or ordinary least-squares regression) is run on the treatment and control groups. The dependent variable is an indicator of being in the treatment group; the independent variables are all of the observable characteristics (for example, location, income level, and relative energy usage in different time periods) along with the relevant hourly pre-treatment loads being examined. For the probit or logit models, the likelihood ratio test of the joint significance of the coefficients should be insignificant. Similarly, for the linear model, the F-test of the significance of the coefficients should be insignificant.

2.4.2 Unbalanced Loads in RCT or RED

If significant differences are found between treatment and control groups in an RCT or RED, then we suggest the following. First, determine whether the differences are likely to have arisen by chance. This can be determined through a combination of statistical analyses of the differences, in light of the size of each group, and an examination of the process through which customers were assigned to groups.

If evidence arises that the sample selection process was corrupted, then the corruption should be corrected when possible. For example, perhaps some customers were wrongly shifted from the control group to the treatment group, but the original assignments can be recovered. If the corruption cannot be corrected, then we suggest using the non-experimental methods we discuss—matching or within-subjects methods, with the understanding that those methods rely on much stronger assumptions and therefore are more likely to produce biased results.

If no evidence arises that the sample selection process was corrupted, but differences between the treatment and control group appear large, then the situation is complicated. No finite sample will ever have perfect balance between treatment and control, but closer balance suggests a more plausible counterfactual. There is no cut-and-dried answer to how big the imbalance must be before it is worth altering an analysis, but if the imbalance seems large enough to have a significant impact on the results, then we suggest that first impact estimates be developed and reported using the methods we describe for RCTs or REDs. We also suggest using an ex post matching procedure to increase balance between the treatment and control groups through re-weighting or truncating the control group, the treatment group or both. This ex post weighting cannot introduce self-selection bias, but it can alter the external validity of the impacts being measured since the impact estimates are now representative of the weighted treatment group. We then suggest using the RCT or RED methods on these groups to produce and report impact estimates that can be compared to the unmatched results to see how large the difference is due to the imbalance.

3. Load Impact Analysis

A useful place to begin an assessment of any pilot that includes event- and/or non-event-based treatments is with a comprehensive examination of the load impacts. These are raw (physical) measurements made to estimate the effects of time-based rates, enabling technology, and various other treatments on customers' levels and patterns of electricity usage. The term load impact, as used in the assessment of pilot treatments, can be defined in several ways, depending somewhat on the nature of the treatments and the overall objectives of the pilot. For example, load impacts are often defined by the reduction in measures of electricity usage (such as a percentage of average annual, daily, or hourly electricity usage) or by the reduction in usage during certain hours of the day, during the system coincident peak demand, or during certain hours of particular days (such as the event hours on days when CPP events are called).

The study of load impacts falls within a broad group of analyses that are primarily statistical and that measure the impacts or effects of the behavior. If one is to have confidence in the estimates of the behavioral parameters derived from the estimated economic models, it is first necessary to be confident in understanding the results from performing the statistical analyses of these load impacts.

As detailed in Section 2, the process of selecting and validating the control group and reference load leads to one of three outcomes: 1) the reference load is from an RCT, 2) the reference load is from a RED, and 3) despite the best efforts to select an appropriate control group, problems in implementation preclude its use as a model to define a reference load based on the experimental design. The situation is similar if, for example, smart meter interval data are available only for customers exposed to the treatment. Under these circumstances, a reference load must be constructed using matching or within-subjects methods.

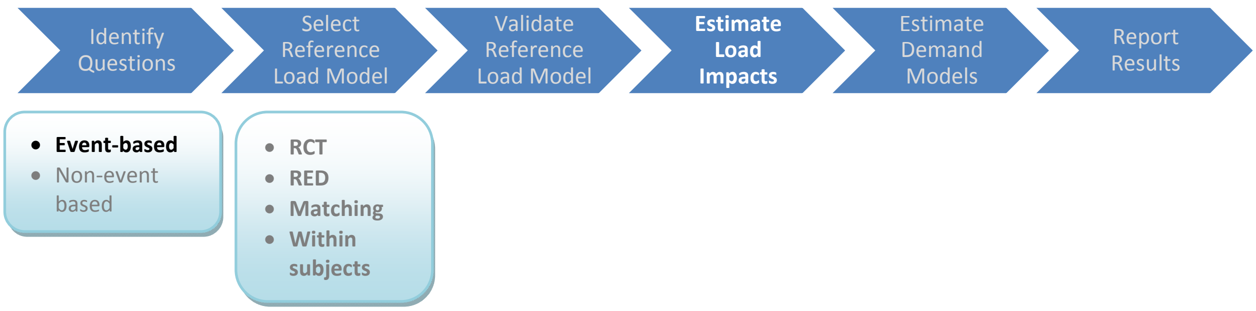
Several statistical methods are frequently used in estimating these load impacts, including fixed-effects estimators, analysis of variance and covariance (ANOVA) and (ANCOVA), difference-in-differences (D-in-D) models, and general treatment effects models. These methods are discussed in detail in Appendices A, B, and C. Fixed-effects regression formulations have been particularly convenient as analysts of recent pilot studies have sought to exploit the panel nature of the load data normally generated through these pilots. The use of a fixed-effects estimator is likely to improve the fit of a regression model because it includes a way to identify a fixed effect for each household. Importantly, this fixed-effects estimator is equivalent to ordinary least-squares regression (OLS) with a dummy variable included for each household.

Both D-in-D and ANOVA/ANCOVA methods can be formulated as fixed-effects regressions. Therefore, in the examples discussed in this section, the focus is on a fixed-effects framework with no loss of generality about the implications.

Note that the suggested analyses are not the only valid ways to estimate load impacts. Other specifications that focus on different levels of aggregation across time or participants may be examined. The analyses shown are relatively straightforward and can be conducted for any CBS pilot. The results should be of interest to those investigating similar types of treatments and should be comparable across pilots.

As mentioned, there are three separate outcomes in the selection of a reference load. The models and methods needed to estimate load impacts from either event-based or non-event based treatments differ depending on which of these three outcomes is obtained. This section describes these methods in a systematic fashion, focusing first on estimating load impacts for event-based treatments and then for non-event-based treatments.

3.1 Event-based Treatments

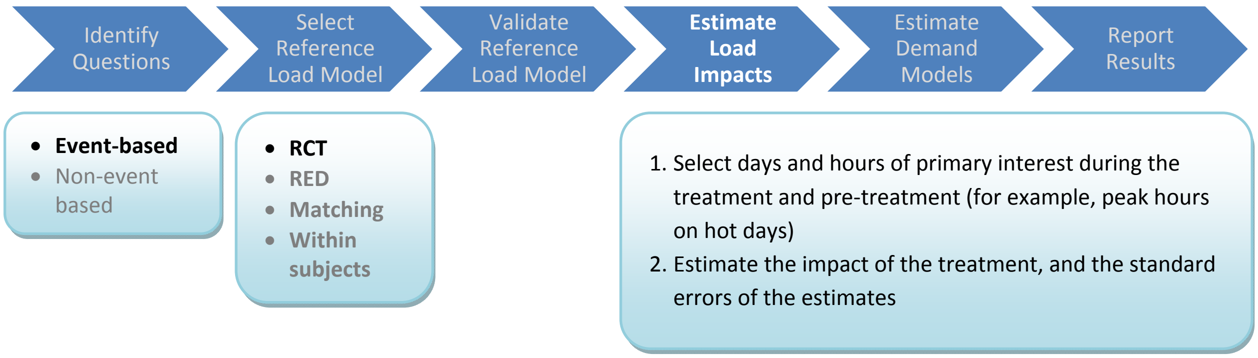


As discussed previously, there are three main types of event-based treatments: CPP, CPR, and utility load control programs. In each case, the utility determines a set number of hours in which customers will either be given a large price incentive to reduce load (CPP and CPR) or will have their load reduced directly for them by the utility (load control). With CPP and CPR, customers are usually notified the day before the event occurs; with load control, customers may or may not be notified. Each type of program focuses on the reduction of usage during peak times when wholesale electricity is expensive or in cases in which system conditions require load reductions. Although these programs may produce net energy savings over the course of a day, week, or billing cycle (the so-called *conservation effect*), they function primarily as a way to reduce peak demand. We focus on an analysis of that aspect of the treatment and, for convenience, all event treatments are generically referred to as a *rate*.

Events are usually focused on very hot summer afternoons, although some utilities may use these programs to reduce peak demand in the winter as well. Much of the following discussion presumes that the focus is on estimating the effects for hot summer days. The same discussion can apply to winter days by focusing on very cold days rather than very hot days. A reference load model for this type of program must consist of a method for estimating what the average usage of the treatment population would have been during the event period(s) had it not been exposed to the treatment—the counterfactual. We focus on the task of estimating the average impact of each event separately.[17]

[17] The same methods (with small variations) can be used for different levels of possible granularity, such as estimating the impact of each hour of each event or estimating the degree of pre- or post-event load shifting. It is also possible to structure the model to estimate the average impact over all events. However, by examining the rate treatments separately, the regression model is algebraically equivalent to one in which multiple treatments are examined. This ability to examine multiple treatments may be particularly helpful, for example, where in addition to a CPP rate treatment, the pilot also includes CPR and load control treatments.

3.1.1 Estimation Using an RCT



Assuming that the reference load is defined by an RCT and that some pre-treatment interval data are available for the entire population of interest, the best method to estimate event load impacts in this context is a form of D-in-D regression. The following example illustrates the method.

Suppose we have the set of days described in Table 4, which lists two hot pre-treatment days and two CPP event days along with relevant data for each. For this example, suppose that hourly load data are available for all customers in the treatment and control groups for May through October. Furthermore, suppose that there were only two CPP events, covering the hours 1–5 PM, and that the four days shown are the hottest of the summer, with the next hottest day having a high temperature of 86°F (30°C).

Table 4: Hypothetical situation for modeling the effect of a CPP rate using an RCT

Date	Day Type	Avg. Load 1-5 PM Control (kW)	Avg. Load 1-5 PM Treatment (kW)	High Temperature (°F)
6/1/2012	Pre-treatment	1.05	1.11	99
6/15/2012	Pre-treatment	1.10	1.19	94
7/9/2012	CPP event	1.43	1.05	100
8/1/2012	CPP event	2.20	1.62	103

The determination of exactly which pre-treatment data to include in the model requires judgment about which, if any, pre-treatment periods occurred under conditions similar to the event periods. This case is fairly typical in that there is a small number of event days and a small number of useful pre-treatment days. Days with temperatures more than 5–10 degrees lower than the event days are not considered useful to adjust for pre-existing differences between groups. Moreover, in this example, it is likely that the useful pre-treatment days are fairly different from the event days because they are earlier in the summer season (for example, June vs. July/August); temperatures are somewhat lower than on the event days. Despite these differences, the pre-treatment days are worth including in the model because they show that there is a noticeable difference in average loads between the treatment and control groups on non-treatment hot days.

Because electricity usage for a household depends on many factors that are consistent over the course of the summer, it is fairly likely that these differences between groups are persistent. If this is true, the estimate of the effect of the treatment can be improved by eliminating from the model these differences in average loads so that the treatment effect is isolated. In expectation, this reduces the model's variance. Once we see what the pre-treatment differences actually are, this step can be viewed as directly subtracting our best estimate of the bias because of imbalance between the groups. The most useful specification would include a set of pre-treatment days with the same conditions as the event days, but this situation is unlikely to occur. Regardless, it is generally better to use pre-treatment data than not to use it.

The recommended regression specification for the example in the table is shown in Equation 3-1:

Eq 3: $load_{it} = a_i + b_1 T_i x I_1 + b_2 T_i x I_2 + b_3 I_1 + b_4 I_2 + u_{it}$

Table 5: Variables in regression Eq. 3

Variable	Descriptions
$load_{it}$	Load in kW for customer *i* at time *t*
a_i	Estimated customer-specific additive constant (frequently referred to as a *customer fixed effects*)[18]
b_1	Estimated average effect of first CPP event
b_2	Estimated average effect of secondCPP event
b_3 , b_4	Estimated effect of each CPP time period on treatment group and control group customers
I_1, I_2	Indicator variables equal to one during the first and second CPP events, respectively; equal to one for all customers during those events
T_i	Indicator variable equal to one for treatment group customers; zero otherwise
$T_i x I_1$	Indicator variables equal to one for treatment group customers during the first and second CPP events, respectively; zero otherwise
u_{it}	The error of the regression for customer *i* at time *t,* which is likely to be correlated over time within customers

The simplest way to represent time (time *t*) is to define the dependent variable as load values equal to the average load for each customer during the event hours on each date (for example, 1–5 PM for each event day), although other specifications of time can also work.

Standard errors for the treatment effects should also be estimated using the "cluster" option available as part of the regression function in most statistical packages. By applying this cluster option at the customer level, the estimated standard errors of the treatment effects will account for correlation in errors at the customer level over time. Failure to use this option can bias the standard error of the treatment effects downward, indicating that the confidence interval around the estimated treatment effect is too narrow and leading to estimates that appear to be more accurate than they actually are. In other words, the level of the estimated parameters is not affected, but the variance of the estimates is biased.

[18] These customer fixed effects are normally incorporated into the model using a fixed effects estimator, as discussed in Appendix A. This fixed effects estimator or its equivalent is contained in many widely used econometrics software packages. In estimation, many apply procedures such as generalized least squares (GLS) to control for heteroskedasticity.

One way to check for errors in analysis or possible underlying inconsistencies in the definition of the variables or the dataset is to compare the results from the recommended regression to the results of a D-in-D calculation, which can be done based on the information in Table 5. Based on Table 4, a D-in-D calculation of b_1 is shown in Eq. 4:

Eq. 4: $b_1 = (1.43 - 1.05) - \left(\frac{1.05+1.10-1.11-1.19}{2}\right) = 0.45$

The first term represents the difference between the treatment and control groups during the first CPP event; the second term represents the average difference between the two groups during the pre-treatment periods.[19] A simple estimate of the standard error of the estimate is equal to the square root of the sum of the squares of the standard errors of the average difference between the pre-treatment and event load in the treatment and control groups. Similarly, a D-in-D calculation of b_2 equals 0.65.

Because it aggregates observations over time, this method (which is, in effect, ANOVA) of calculating b_1 and b_2 is likely to result in an estimate that is less precise than the value generated from the panel data regression model in Eq. 4.[20] We are not aware of an analogous secondary method for checking the regression standard errors.

The recommended regression in Eq. 4 is based on the assumption that pre-treatment data are available, which may not always be true. In this case, it is impossible to adjust for pre-treatment differences between treatment and control groups, but it is possible to estimate load impacts using load for the control group as a reference load. This will result in a treatment effect estimate that is still unbiased but that may be less precise. The appropriate regression specification in this case is shown in Eq.5:

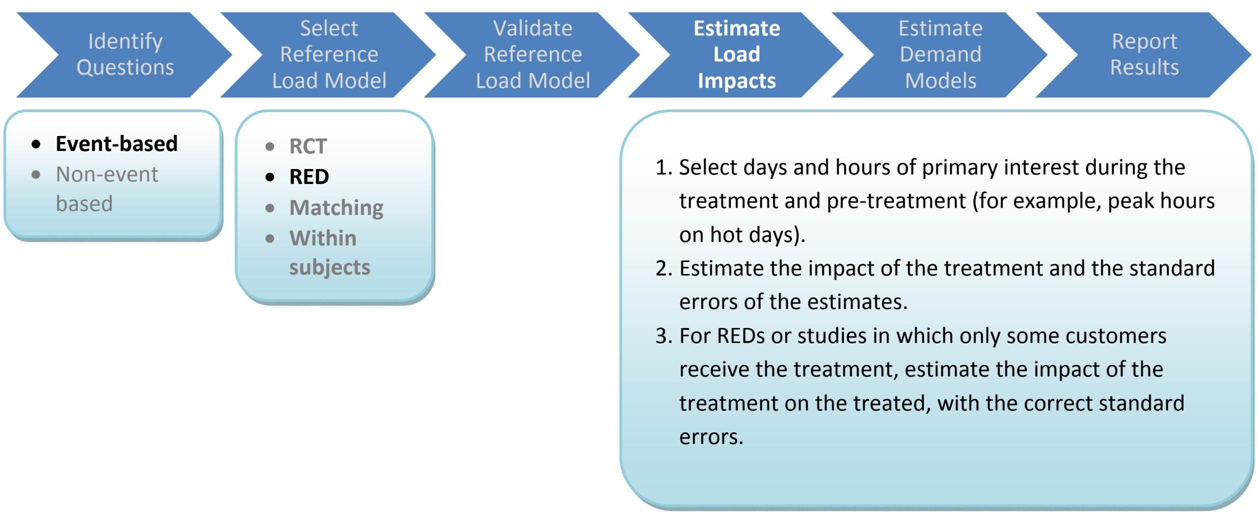
Eq. 5: $load_{it} = a + b_1 T_i x I_1 + b_2 T_i x I_2 + u_{it}$

Note that here there are no customer-specific constants as there were in Eq.3. [21] In this case, a way to check for analysis errors is to compare the results from the recommended regression to the results of subtracting the average load in the treatment group during the event from the average load in the control group during the event, which could be done with the last two rows of Table 4Table 3-1 (that is, excluding the pre-treatment numbers). The standard error of the estimate is equal to the square root of the sum of the squares of the standard errors of the average load in the treatment and control groups—the typical calculation for estimating standard errors of differences.

[19] This particular way to estimate b_1 is equivalent to the D-in-D estimator given in Equation 3-2, with the exception that the event day effects are estimated separately and averaged in the calculation of b_1.
[20] It must be remembered that if the regression function in Equation 3-2 is estimated by OLS rather than the fixed-effects estimator, the results will be equivalent to those in the decomposition table in Appendix A.
[21] This is a direct result of the fact that in order to include customer fixed effects in these models estimated by a fixed-effects estimator, any variable that does not differ over time must be included as an interaction term with a variable that differs over time. Without the pre-treatment data, the variables $t_i x I_1$ and $t_i x I_2$ are constant for each customer over time, so if they are included in the regression (as they must be), the "fixed effects" cannot be used. See Appendix A for a detailed discussion.

One issue that has to be addressed is that most utility billing is done on monthly cycles with bill dates that differ across customers. This means that some approximation has to be used to assign customer usage values to particular months. A simple method is to assign to a particular month all bills that were finalized prior to the middle of the next month and to assign to the next month all bills that were finalized after that. The month during which each customer was enrolled in the treatment could be excluded from the analysis because it would constitute a partial treatment month. A more complicated method that may be more accurate is to assign each day a usage value based on the monthly bill divided by the number of days in that billing cycle. Then, each daily value can be assigned to the month in which it occurs. For example, if a customer used 600 kWh in a billing cycle from April 15–May 14, the daily value for each of those days would be 600 kWh/30 days = 20 kWh/day. That value would then be assigned to the last half of the days in April and the first half of the days in May. Again, the month that each customer enrolled on the treatment could be excluded.

3.1.2 Estimation Using an RED



We use the same example to illustrate the methods to estimate event-day load impacts when the reference load is defined using an RED. Under this RED method, a population of customers is still subject to random assignment to treatment and control groups, but the treatment consists of an encouragement to take up the rate (as described in Section 2). The encouragement induces a correlation between being in the treatment group and being on the rate, which can then be used to produce an unbiased estimate of the effect of the rate on the customers who took up the rate because of the encouragement (called a *local average treatment effect*, or LATE). Theoretically, even weak encouragement could work for estimating load impacts if sample sizes were large enough. However, as a practical matter, it must be the case that 1) a large fraction of customers who receive encouragement actually take up the treatment and 2) only a small fraction of customers who do not receive encouragement learn of the opportunity and take up the treatment on their own (and, in a special case of REDs, control customers who do not receive the encouragement are in fact not allowed to take up the treatment if they find it on their own). For this reason, it is important to report these data up front, as is done in Table 6. The table shows a (hypothetical) reasonably successful RED in which almost two-thirds of encouraged customers took the offered rate and a very small fraction of non-encouraged customers did so.[22] In the case of an RCT in which treatment customers are allowed to opt-out, the encouraged group consists of the treatment group, and the fraction accepting the rate is the fraction that chooses not to opt-out.

Table 6: Percentage of customers in the encouraged and not encouraged groups accepting the rate

Group	Accept Rate (%)	Refuse Rate (%)
Encouraged	65	35
Not encouraged	2	98

The most straightforward method for estimating load impacts in this case is to perform two analysis steps: first, an estimate of the ITT effect of the encouragement and second, a scaling exercise to estimate the LATE. To estimate the ITT effect, use the same regression function described for RCTs. In this case, the RCT regression (Eq. 3) is altered so that $T_i xI_1$ and $T_i xI_2$ are equal to unity for any customer in the encouraged group during Critical Peak Events 1 and 2, respectively, and zero otherwise. This means that the estimated coefficients on those variables represent the effects of a CPP event on customers encouraged to sign up for the rate rather than the effect of a CPP event on customers who actually signed up for the rate.

Second, to estimate the LATE, or the effect of the treatment on customers who actually took up the treatment because they were encouraged (rather than the effect of only the encouragement), the estimated ITT effect for each event (that is, b_1 or b_2 from Eq. 3) is divided by the difference in the fraction of customers who took up the rate between the encouraged group and the non-encouraged group:

$$LATE = \frac{ITT}{\%\ of\ encouraged\ that\ signed\ up - \%\ of\ not\ encouraged\ that\ signed\ up}$$

In Table 6, for example, that fraction equals 0.65 - 0.02 = 0.63.[23] This is equivalent to scaling up the ITT encouragement effect for each event by 1/0.63 = 1.59. Note that this fraction might be different for each CPP event if customers decide to sign up for or drop out of the rate between events.

The standard error of the LATE estimate can be developed similarly. To do so, we take the standard error of the ITT effect estimate and divide it in the same way as the estimate: by the difference in the fraction of customers who took up the rate between the encouraged group and the non-encouraged group. This calculation illustrates the way in which lower take-up levels in the encouraged group lead to inflated standard errors of the treatment effect.

$$SE_{LATE} = \frac{SE_{ITT}}{\%\ of\ encouraged\ that\ signed\ up - \%\ of\ not\ encouraged\ that\ signed\ up}$$

[22] Note that the numbers in this example are closer to what would be expected from a default (or opt-out) encouragement. Opt-in recruiting to a rate would probably produce much lower subscription rates.

[23] Note that in the special case of an RED in which the control customers (the non-encouraged customers) are not allowed to sign up for the treatment, LATE is simply the estimate of the ITT effect divided by the fraction of customers who signed up for the program out of all of those encouraged to do so.

As an alternative to this analysis of first estimating the ITT effect and then scaling it to get an estimate of LATE, LATE can be estimated directly using a two-stage instrumental variables approach. Taking as an example the hypothetical CPP situation from Table 4, the first-stage regressions have as their dependent variables indicators for being on the rate during each CPP event and as their independent variables a constant—an indicator for being in the encouraged group and indicator variables for each CPP period. Note that the first stage is estimated over all time periods, including pre-treatment time periods. The dependent variables in the first-stage regressions will equal zero for all customers during all pre-treatment time periods. For treatment time periods, there are two variables to be instrumented: being on the CPP rate during the first CPP event and being on the CPP rate during the second CPP event. Consequently, we need at least two instruments. In this case, there are exactly two instruments: being in the encouraged group during the first CPP event and being in the encouraged group on the CPP rate during the second CPP event.

The second stage is then the same regression specification as in Eq. 3, but the predicted values from the first stage are used instead of $T_i x I_1$ and $T_i x I_2$. This system of equations is shown below:

Eq. 6:

$$T_i x I_1 = a_{1i} + b_5 I_E x I_1 + b_6 I_E x I_2 + b_7 I_1 + b_8 I_2 + \varepsilon_{it}$$
$$T_i x I_2 = a_{2i} + b_9 I_E x I_1 + b_{10} I_E x I_2 + b_{11} I_1 + b_{12} I_2 + \epsilon_{it}$$
$$load_{it} = a_i + b_1 \widehat{T_i x} I_1 + b_2 \widehat{T_i x} I_2 + b_3 I_1 + b_4 I_2 + u_{it}$$

This two-stage process may have to be completed more than once if the fraction of customers on the rate changed significantly between events. In that case, the same specification holds—but only events in which the fraction of customers on the rate stayed nearly constant should be included in the same estimation.

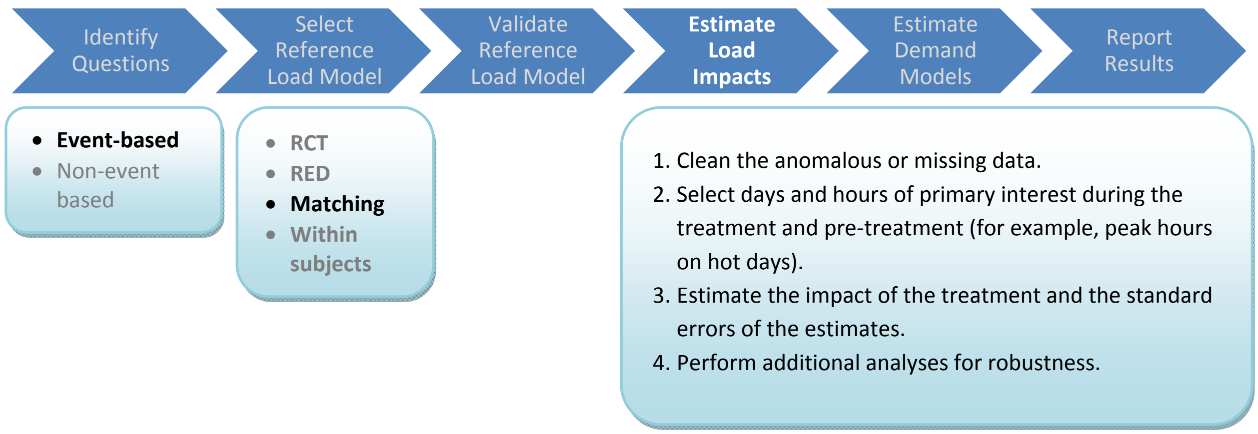
The two-stage instrumental variables process should produce estimates of LATE identical to those obtained through the first method: estimating the ITT effect and scaling it by dividing by the difference in the take-up rate between encouraged and non-encouraged customers.

Table 7: Variables in regression Eq. 6

Variable	Descriptions
$load_{it}$	Load in kW for customer *i* at time *t*
a_i, a_{1i}, a_{2i}	Estimated customer-specific additive constant
b_1	Estimated average effect of first CPP event who are on CPP
b_2	Estimated average effect of second CPP event who are on CPP
b_3 , b_4	Estimated effect of each CPP time period on treatment group and control group customers
b_5, b_6, b_9, b_{10}	Estimated effect of the encouragement on CPP take-up during each CPP event; note that b_6 and b_9 should be very close to zero
b_7, b_8, b_{11}, b_{12}	Estimated effect of CPP time periods in the first-stage regressions
I_1, I_2	Indicator variables equal to one during the first and second CPP events, respectively; equal to one for all customers during those events
I_E	Indicator variable equal to one for customers in the encouraged group; zero

	otherwise
T_i	Indicator variable equal to one for treatment group customers; zero otherwise
$\widehat{T_\iota x I_1}, \widehat{T_\iota x I_2}$	Fitted values from the first-stage regressions
u_{it}	The error of the regression for customer *i* at time *t,* which is likely to be correlated over time within customers
$\boldsymbol{\varepsilon_{it}}, \boldsymbol{\epsilon_{it}}$	Error terms in the first-stage regressions

3.1.3 Estimation Using a Matched Control Group



If an RCT or RED was not implemented or was implemented unsuccessfully, there is no control group to serve as the reference load. One possible remedy is to identify a matched control group and then perform the same D-in-D regression analysis described in the sub-section on using an RCT for estimation. As mentioned previously, this method is inferior to the RCT or RED designs for two reasons:

There may be a very limited set of observable variables to use for matching customers. This means that there will probably be noticeable biases between the groups, regardless of the matching procedure used.

The use of a matched control group relies on the unconfoundedness assumption; we must assume that there are no unobservable variables correlated with both being on the treatment and usage behavior.

Regardless of how well the matching procedure appears to work, we can never rule out the possibility of unobservable differences between groups producing bias in the results. In the case of event-based programs in which the underlying rate in the treatment group is the customer's default rate, if only small differences between the treatment group and the matched control group are observed on non-event days that are similar to event days, one may be more convinced that the control group is a good reference case and that therefore the bias is small. However, this is always a subjective determination: all coefficient estimates and judgments of statistical significance become conditional on the unconfoundedness assumption, which is unknowable (unless one is specifically testing the level of bias by running an RCT).

The matching process uses propensity score matching to identify customers who are not subject to the treatment but who have similar characteristics to customers who are subject to the treatment. Several similar methods can be used to control for selection (described in EPRI report 1020855). For the method described here, the matching process leads to resulting (matched) control and treatment groups that can then be used in an analysis identical to that for an RCT. The matching process attempts to use observable variables such as those listed next to predict how likely customers are to participate. Customers who did not choose to participate but who are predicted to have a somewhat high propensity to participate are then used as the matched control group.

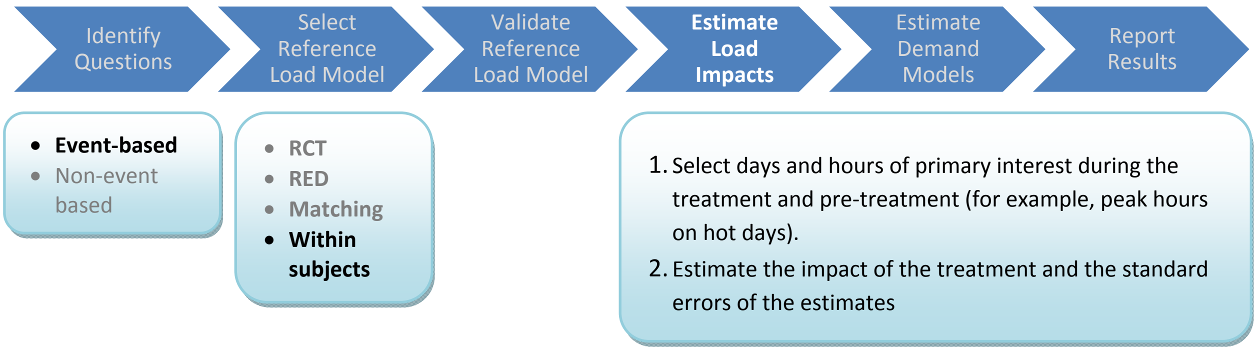
For a study with event-based treatments, propensity score matching is based on the following observable variables in order of priority, with the understanding that not all of them will be available:

1. Hourly usage at the customer level during the afternoon and evening on hot, pre-treatment days
2. If pre-treatment data are not available, hourly usage at the customer level during the afternoon and evening on hot, non-event, post-treatment days if the event-based treatment is the only treatment the customer is subject to (that is, the underlying rate is not TOU, and the customer has not been provided with an IHD or PCT)
3. Customer-level demographic and location information, such as size of house, income, number of people in the household, age, and ZIP code
4. As a last resort, aggregate-level demographic information such as income, size of house, and age, based on census block group data

Having performed the propensity score match to produce a control group, the regression analysis to measure load impacts is the same as that described in the discussion of RCT methods. However, again we emphasize that the interpretation of results after a matching exercise is not the same as in the RCT case. Standard measures of statistical significance are conditional on the unconfoundedness assumption, which is unknowable. In addition to the method described here, other methods exist for controlling for selection bias.[24]

[24] For example, the propensity score, rather than being used to create the control group, may be added as a variable in part of the regression. See Imbens and Wooldridge (2009) for this and other examples.

3.1.4 Within Subjects Methods



It may not be possible to use treatment-control methods for several reasons. For example, it may be the case that smart meter interval data are required to measure and test impacts, and they are available only for customers exposed to the treatment. In the case of event-based treatments, a reference load model and an estimate of the effect of the treatment can still be produced using measured load of the subjects in the treatment group during non-event periods. However, within-subjects methods require more assumptions than matching methods (as described in Section 2) and may therefore be more susceptible to bias because of inaccurate assumptions. These methods should be used only if other methods are not possible.

The estimated reference load for an event in this case represents an estimate of what the load would have been in the treatment group had an event not been called. The validity of the within-subjects method relies on two assumptions. First, it assumes that there is a set of non-event days that has sufficiently comparable characteristics to the set of event days—so that it is reasonable to assume that the only difference in customer behavior during event days and the comparable non-event days is that they experienced the treatment in effect during event days (for example, the comparable non-event days have similar temperatures and daylight hours to those of the event days). Second, if the underlying rate is the default rate in the population, the assumption is that customers do not alter their non-event behavior very much as a result of being on the rate. If both of these conditions are fulfilled, non-treatment days are an acceptable approximation of what the load would have been if the treatment group had never been on the rate. This may frequently be the case.

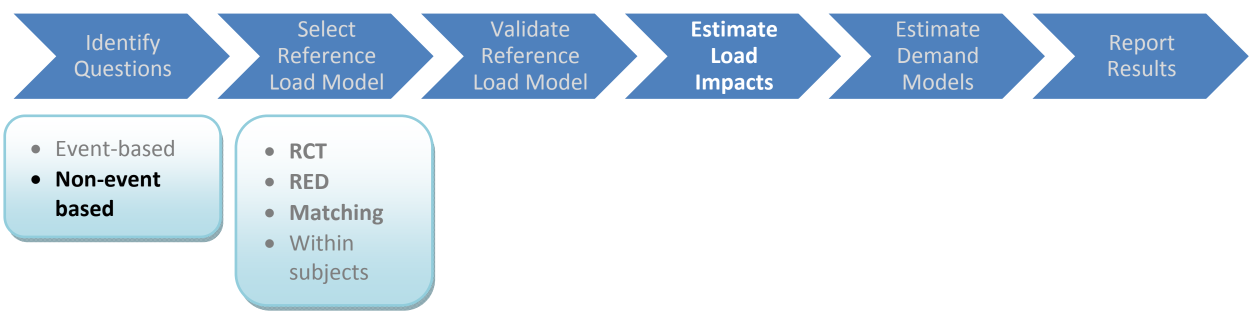
Within-subjects models consist of either panel regressions on only the treatment group or regressions run individually for each customer. In each case, event impacts are measured using indicator variables for event periods. Although we will not go into the details of this method here (there are many industry examples of these types of analysis),[25] the following suggestions are provided:

---

[25] For an example of a within-subjects method, see many of the demand response evaluations published in California over the past several years. For a particular example, see "2010 Load Impact Evaluation of San Diego Gas & Electric Company's Summer Saver Program" prepared by the FSC Group and available at: http://www.calmac.org/publications/SDGE_2010_Summer_Saver_Load_Impact_Report.pdf Also see EPRI report 1023644.

- There is little value to including in the data set non-event days in which temperatures are far different than event temperatures. Given that CPP days are often the hottest days of the summer, a typical problem in this modeling approach is a lack of relevant non-event days for modeling.
- In determining which variables to include and which conditions to attempt to control for, it is best to keep expectations modest. The amount of independent variation in weather at different times of day over the course of a summer is usually low. The same is true of variation in seasonal conditions and day-of-week conditions independent of weather.=
- The predictive accuracy of any model can be assessed using an out-of-sample testing regime[26] in which several non-event days with event-like conditions are withheld from the model during fitting. Load predictions from the model can then be compared to actual load on these days to assess predictive accuracy. The result of this exercise should be displayed in graphs. This exercise also limits the potential for over-fitting the model (that is, adding meaningless variables). Over-fitting to the out-of-sample days is still possible. Whether a model is being over-fit requires some judgment. The question for any given variable's inclusion is this: Is it really plausible that there is enough variation in this variable, independent of all of the other variables, for this coefficient to be well-measured? For example, if we think that morning temperature in particular is a useful predictor of CPP impacts in addition to daily average temperature, we need to determine whether the data provide significant variation in morning temperature for the relevant range of daily average temperatures.

3.2 Non-event Based Treatments



As indicated, there are several types of non-event treatments: TOU rates, IHDs, PCTs, some kinds of feedback information, educational efforts, and combinations of them are referred to generically as *treatments* in this section.[27] Customers are exposed to a treatment upon the

[26] Also commonly referred to as *cross-validation*.
[27] Some feedback mechanisms involve providing customers with a single dose of information designed to help them better understand when and how they use electricity; this single dose qualifies as a non-event treatment. Others involve providing routine information, which also qualifies as a non-event treatment. Some provide only periodic

initiation of the treatment period that provides a motivation to change usage behavior—through incentives, by providing more information about usage behavior, or by providing tools that more easily allow changes in usage behavior. Furthermore, the treatment is always present after its onset rather than only during certain critical events. For these types of treatments, it might be reasonable to hypothesize that in addition to leading to a reduction in usage during certain targeted times of the day, the treatment might also lead to an overall reduction in energy use and to a shift in usage from targeted to non-targeted times of day. We discuss the analysis of both of these effects.

In addition to the fact that these treatments are always present, there is another important difference between the event- and non-event-based treatments that directly affect the ability to measure impacts of non-event-based treatments. The effect of non-event-based treatments on usage at any given time is likely to be smaller than that of an event-based treatment during an event, in part because what is being measured is the annual change in electricity usage—not usage for a few hours of a day. As a practical matter, this means that there are more stringent analytical requirements for measuring the effect of non-event-based treatments well enough to establish their value. Therefore, a larger sample size may be needed to achieve statistically significant estimates. If the price signal (the peak to off-peak price ratio) in TOU rate is quite low (3:1 or less), the expected effect on peak period usage is usually 5% or less. The effects of IHDs, PCTs, or education are now being tested through pilots and experiments. Early indications suggest that total electricity usage changes are less than 5%. The energy conservation effects of these programs are also likely to be modest—in the range of 0% to 5% of usage. A synthesis of pilot results is available in EPRI report 1025856 (2013).

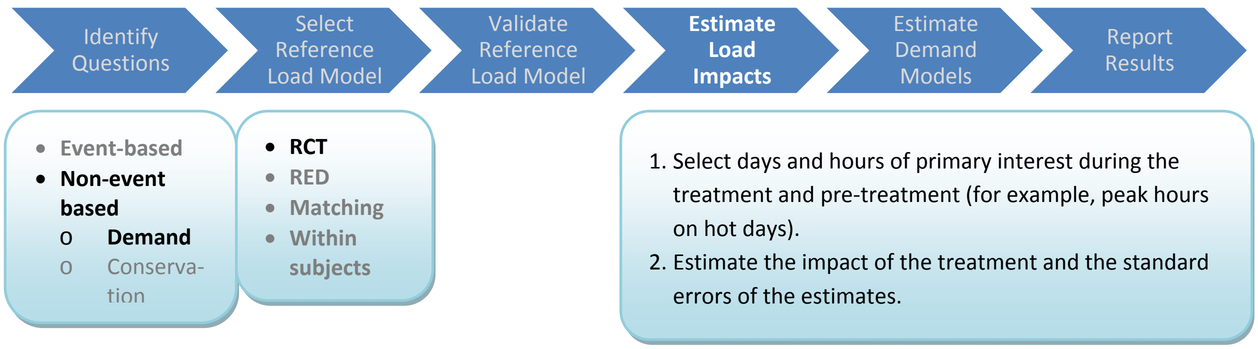
Given these characteristics, it is unlikely that practically useful estimates of the effects of these treatments can be developed using any design other than an RCT or a RED with a low rate of non-compliers. As described in Section 2, the assumptions required for using matching or within-subjects methods are likely to be less valid with non-event-based treatments. For that reason, the discussion is limited to estimating energy use effects and shifting effects based on these two experimental designs.

We discuss the estimation of two important load impacts for these non-event-based treatments: demand-shifting and energy conservation. *Demand-shifting* in this context refers to any change in usage during particular hours, regardless of whether that usage change is made up for during other hours (this may be less likely if the treatment is designed to elicit conservation). This may or may not be the case, but an estimation scheme that covers all hours of the day can answer that question.

---

information that would qualify as a sort of event-based treatment, except that the timing of the treatment is not associated with an extreme day of system conditions (EPRI report 1020855).

3.2.1 Estimation of Demand-Shifting Using an RCT



Estimating the demand-shifting effect of these rates based on an RCT is best done using a D-in-D regression, similar to the one for event-based programs using Eq. 3. It may be of interest to measure usage shifting behavior for any particular hour of the day or for a particular block of hours. Consider the example in which the interest is estimating the degree to which a treatment causes customers to shift usage away from the time period of 1–5 PM on weekdays.

We recommend estimating an effect that is specific to the days and/or months of interest and separating the analysis as such. For example, if the target of the treatment is reduction during weekdays, weekdays should be separated from weekends; likewise, if summer is the target, it should be separated from other seasons. As a secondary analysis, a further separation of days (or an inclusion of interaction variables for those categories) may provide insight into the effectiveness of the treatment. Specifically, dividing days into at least two categories based on weather conditions may be useful if customers may have more scope to shift load at times when it is very hot or very cold. Another possibility is to explicitly model effects of weather in the model. However, this requires specifying a functional form for weather, which both increases the complexity of the model and may lead to misspecification.

In this example, the recommended fixed-effect regression to estimate the effect of the treatment under normal conditions during 1–5 PM on weekdays in the summer is shown in Eq. 7:

Eq. 7: $load_{it} = a_i + b_1 T_i x I_1 + b_2 I_1 + u_{it}$

The load data that should be included in the regression are loads measured at the customer level during the periods of interest in that specific regression: in this example, summer weekdays from 1 to 5 PM. Both the treatment and pre-treatment periods should be included. Separate regressions can be run for different combinations of day type, season, and conditions. Similarly, the regressions can be run for any time of day to determine the average impact at that time (for example, the regression could also be run for every hour except for 1–5 PM to estimate the degree to which usage shifted from peak to off-peak).[28]

[28] It is also possible to use interactions of indicator variables to produce these results all at once. However, because it is sufficient to use separate regressions, this is not discussed further.

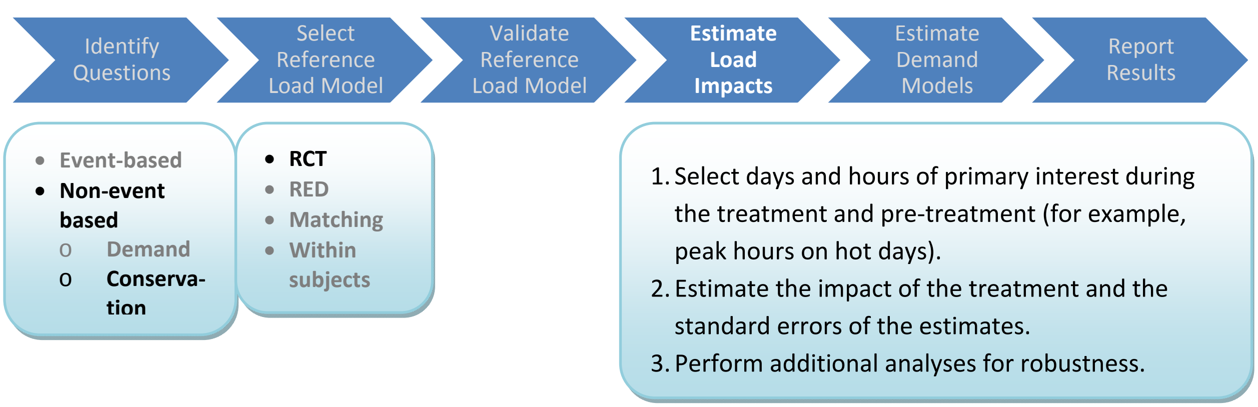
Table 8: Variables in regression Eq.7

Variable	Descriptions
$load_{it}$	Load in kWh for customer *i* at time *t*, including only weekdays in the summer under normal conditions
a_i	Estimated customer-specific additive constant
b_1	Estimated average effect of the treatment
b_2	Estimated effect of the treatment period for all treatment group and control group customers
I_1	Indicator variable equal to one during treatment periods; zero otherwise
$T_i x I_1$	Indicator variables equal to one for treatment group customers during the treatment period, respectively; zero otherwise
u_{it}	The error of the regression for customer *i* at time *t,* which is likely to be correlated over time within customers

As an example of the secondary analysis mentioned previously, suppose that the area of interest is in the Midwest and that all days are divided into categories *normal* and *extreme*, where *normal* is defined as having high temperatures in the range of 25–85°F (4–29°C) and *extreme* is defined as temperatures outside that range. This categorization separates summer days between those that are mild and those that are very hot, and winter days between those that are mild and those that are very cold. The specification in Equation 3-5 could then be used separately for days in each category.

Standard errors of the treatment effect should be estimated using the "cluster" option available in most statistical packages as part of the regression function. Applying clustering at the customer level will produce estimates of standard errors that account for correlation in errors at the customer level over time.

3.2.2 Estimation of Energy Conservation Using an RCT



Estimating energy conservation effects of one of these treatments does not require smart meter interval data; it can be performed using monthly billing data alone (although using smart meter interval data may increase the precision of the estimate). Estimating the overall energy savings associated with such a treatment is done using a specification similar to Eq. 7. Suppose that we are estimating the energy savings associated with an IHD that has been in place for a treatment group for a year. Also suppose that we have one year of pre-treatment monthly billing data for all customers in the treatment group and the control group. To estimate the overall average energy savings associated with the IHD during the year-long experiment, we would regress monthly usage for each customer for each month onto a customer-specific constant—an indicator equal to one during the treatment period and zero otherwise and an indicator equal to one for customers in the treatment group during the treatment period and zero otherwise. This specification is shown in Eq. 8:

Eq. 8: $load_{im} = a_i + b_1 T_i x I_1 + b_2 I_1 + u_{im}$

Table 9: Variables in regression Eq. 8

Variable	Descriptions
$load_{im}$	Monthly usage in kWh for customer *i* during month *m,* including only weekdays in the summer under normal conditions
a_i	Estimated customer-specific additive constant
b_1	Estimated average effect of the treatment
b_2	Estimated effect of the treatment period for all treatment group and control group customers
I_1	Indicator variable equal to one during treatment periods; zero otherwise
$T_i x I_1$	Indicator variables equal to one for treatment group customers during the treatment period, respectively; zero otherwise
u_{im}	The error of the regression for customer *i* during month *m,* which is likely to be correlated over time within customers

A common variation on this specification includes indicator variables for each month-year combination rather than a single indicator variable for the treatment period—primarily to test for persistence of the treatment effect, but also to test for seasonal effects.[29] In addition, this specification can be adapted to include different estimated effects for each month.

Standard errors of the treatment effect should be estimated using the "cluster" option available in most statistical packages as part of the regression function. Applying clustering at the customer level will produce estimates of standard errors that account for correlation in errors at the customer level over time.

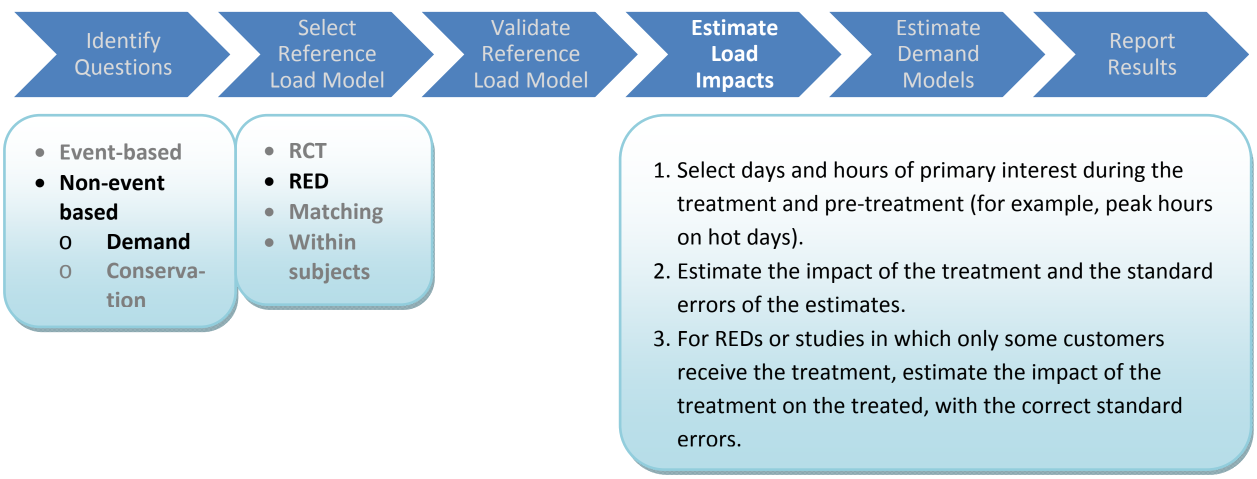
Rather than estimating the model using absolute load values, there might be interest in using the logarithm of load as the dependent variable to directly estimate percentage reductions in energy usage. Although there may be some value in that exercise,[30] we do not recommend it. Instead, we recommend estimating the model on raw load values and then converting the estimated effect to a percentage of observed usage. Estimating the effect using the logarithm of usage tends to equalize the effect of customers with different levels of overall usage. If large customers tend to reduce usage much more than small customers as a fraction of their overall usage, this likely understates the energy conservation effect in the population. Keeping the dependent variable

[29] Alcott, H., "Social Norms and Energy Conservation," *Journal of Public Economics* 95 (2011) 1082–1095.
[30] In particular, the time-based and customer-specific indicator variables tend to explain more of the variation in usage in a model in which load is converted into logarithmic terms rather than in a linear model.

linear avoids this problem. An alternative would be to use the logarithm of usage but to fit separate treatment effects for customers based on their overall level of usage.

One issue that has to be addressed is that most utility billing is done on monthly cycles with bill dates that differ across customers. This means that some approximation has to be used to assign customer usage values to particular months. A simple method is to assign to a particular month all bills that were finalized prior to the middle of the next month and to assign to the next month all bills that were finalized after that. The month during which each customer was enrolled in the treatment could be excluded from the analysis because it would constitute a partial treatment month. A more complicated method that may be more accurate is to assign each day a usage value based on the monthly bill divided by the number of days in that billing cycle. Then, each daily value can be assigned to the month in which it occurs. For example, if a customer used 600 kWh in a billing cycle from April 15–May 14, the daily value for each of those days would be 600 kWh/30 days = 20 kWh/day. That value would then be assigned to the last half of the days in April and the first half of the days in May. Again, the month that each customer enrolled on the treatment could be excluded.

3.2.3 Estimation of Demand Shifting Using an RED



As above, there are two main ways to develop estimates of load impacts and of standard errors for an RED design. The most straightforward method to develop load impacts is to use the same regression function described for RCTs to first estimate the ITT effect of the encouragement. In this case, the RCT regression (Equation 3-5) is altered so that T is equal to one for any customer in the encouraged group during the treatment period. This means that the estimated coefficient on that variable represents the effect of the treatment on customers encouraged to take the treatment rather than on customers who necessarily took the treatment.

The second step is to estimate the LATE, or the effect of the treatment on those who actually took up the treatment (rather than the effect of the encouragement). LATE can be estimated by dividing the effect of the encouragement (that is, b_1 from Equation 3-5) by the difference in the fraction of customers who took up the rate between the encouraged group and the non-encouraged group. For example, considering the example in Table 3-3, that fraction equals 0.65 - 0.02 = 0.63. This is equivalent to scaling up the ITT encouragement effect by 1/0.63 = 1.59. As discussed in the event-based section above, to develop standard errors for the LATE estimator, the standard error estimates of the TOT encouragement effect can also be divided by the fraction of customers who took up the rate in the encouraged group minus the fraction who took up the rate in the non-encouraged group.

As an alternative to scaling the regression results of the encouragement, the regression can be specified in terms of instrumental variables, with the encouragement during the treatment period acting as an instrument for the treatment during the treatment period. Taking as an example the hypothetical situation in Equation 3-5 but supposing that an RED was implemented instead of an RCT, the first-stage regression has as its dependent variable an indicator for having the treatment during the treatment period and as its independent variables a constant: an indicator for being in the encouraged group and an indicator variable for the treatment period. Note that the first stage is estimated over all time periods, including pre-treatment time periods. The dependent variable in the first-stage regression will equal zero for all customers during all pre-treatment time periods. The second stage is then the same regression specification as in Equation 3-5 but using the predicted values from the first stage in place of $\mathrm{T_i x I_1}$. This system of equations is shown below:

Eq. 9: $T_i x I_1 = a_{1i} + b_3 I_E x I_1 + b_4 I_1 + \varepsilon_{it}$
Eq. 10: $load_{it} = a_i + b_1 \widehat{T_i x I_1} + b_2 I_1 + u_{it}$

Table 10: Variables in regression Eq. 9 and Eq. 10

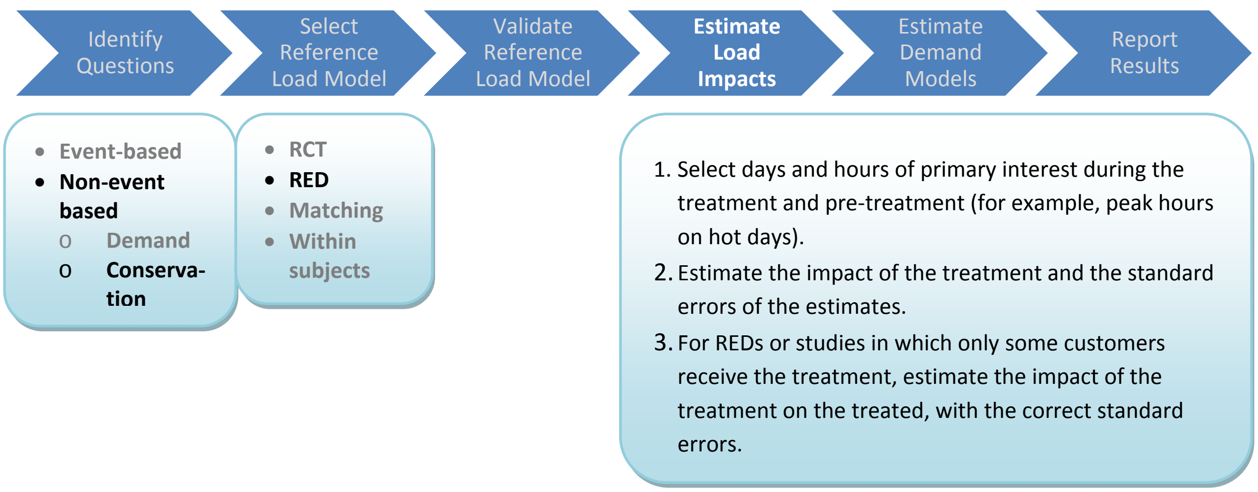
Variable	**Descriptions**
$load_{it}$	Load in kWh for customer *i* at time *t,* including only weekdays in the summer under normal conditions (but here *t* is a specified set of hours: those when demand measurement counts)
a_i, a_{1i}	Estimated customer-specific additive constant
b_1	Estimated average effect of the treatment
b_2	Estimated effect of the treatment period for all customers
b_3	Estimated average effect of the encouragement on treatment take-up
b_4	Estimated effect of the treatment period in the first stage regression
I_1	Indicator variable equal to one during treatment periods; zero otherwise
I_E	Indicator variable equal to one for customers in the encouraged group; zero otherwise
T_i	Indicator variables equal to one for treatment group customers; zero otherwise
$T_i x I_1$	Indicator variables equal to one for treatment group customers during the treatment period, respectively; zero otherwise
$\widehat{\boldsymbol{T_i x I_1}}$	Fitted values from Eq. 9
u_{it}	The error of the regression for customer *i* during time *t,* which is likely to be correlated over time within customers
ε_{it}	The error in the first-stage regressions

If the fraction of customers on the rate in the encouraged and non-encouraged groups changes significantly during the study period, it may be useful to repeat this analysis separately for time periods of relatively stable treatment rates.

Just as in the RCT case, standard errors should be estimated using the “cluster” option available in most statistical packages as part of the regression function. Applying clustering at the customer level will produce estimates that account for correlation in errors at the customer level over time.

The two-stage process should produce treatment effect estimates identical to those obtained through the first method of running an ITT analysis on the encouraged variable and dividing by the difference in the take-up rate between encouraged and non-encouraged customers. The two-stage process will produce slightly different standard errors than the first method, but they are unlikely to be materially different.

3.2.4 Estimation of Energy Conservation Using an RED



The same basic principles discussed above apply to using an RED to estimate energy conservation rather than demand shifting. First, an estimate of the ITT effect can be developed, with a variable for being in the encouraged group used in place of the treatment group variable. To estimate the LATE (the effect of the treatment on customers who took up the treatment because they were encouraged), that estimate can then be scaled by the difference between the fraction of encouraged customers who took up the rate and the fraction of non-encouraged customers who took up the rate.

Again, standard error estimates can be developed by multiplying the standard error estimates for the encouraged variable in the ITT regression on encouragement by the difference in the fraction of customers who took up the rate between the encouraged and the non-encouraged group.

As in “Estimation of Demand-Shifting Using an RED” and extending the example by supposing that the situation in Eq. 8 is implemented as an RED rather than an RCT, the LATE estimate can also be specified by using instrumental variables. The first-stage regression has as its dependent variable an indicator for having the treatment and as its independent variables a constant: an indicator for being in the encouraged group during the treatment period and an indicator equal to one during the treatment period and zero otherwise. The second-stage regression is the same as Eq. 8 but with the fitted values from the first stage used in place of the treatment variable.

Eq. 11: $T_i x I_1 = a_{1i} + b_3 I_E x I_1 + b_4 I_1 + \varepsilon_{im}$
Eq. 12: $load_{im} = a_i + b_1 \widehat{T_\iota x I_1} + b_2 I_1 + u_{im}$

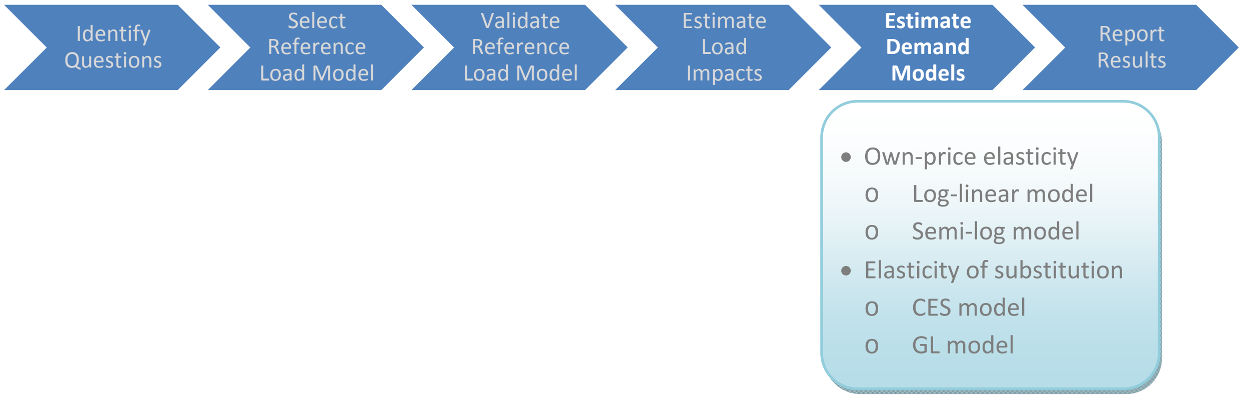
Table 11: Variables in regression Eq. 11 and Eq. 12

Variable	Descriptions
$load_{im}$	Monthly usage in kWh for customer *i* during month *m,* including only weekdays in the summer under normal conditions
$a_{i,} a_{1i}$	Estimated customer-specific additive constant
b_1	Estimated average effect of the treatment
b_2	Estimated effect of the treatment period for all customers
b_3	Estimated effect of the encouragement on treatment take-up
b_4	Estimated effect of the treatment period in the first stage regression
I_1	Indicator variable equal to one during treatment periods; zero otherwise
I_E	Indicator variable equal to one for customers in the encouraged group; zero otherwise
T_i	Indicator variables equal to one for treatment group customers; zero otherwise
$T_i x I_1$	Indicator variables equal to one for treatment group customers during the treatment period, respectively; zero otherwise
$\widehat{\boldsymbol{T_\iota x I_1}}$	Fitted values from Eq. 11
u_{im}	The error of the regression for customer *i* during month *m,* which is likely to be correlated over time within customers
ε_{im}	The error in the first-stage regression for customer *i* during month *m*

As in the case of Eq. 8, a common variation would include indicator variables for each month-year combination. Standard errors should be estimated using the "cluster" option available in most statistical packages as part of the regression function. Applying clustering at the customer level will produce estimates that account for correlation in errors at the customer level over time.

The two-stage process should produce LATE estimates identical to those obtained through the first method of running the ITT regression on the encouraged variable and dividing by the difference in the take-up rate between encouraged and non-encouraged customers.

4. Models of Price Response



The methods previously discussed are used primarily to estimate load impacts, often defined by the reduction in measures of electricity usage such as a percentage of average annual, daily, or hourly electricity usage, or by the reduction in usage during certain hours of the day or during the system coincident peak demand. These physical measurements estimate the effects of time-based rates, enabling technology, and various other treatments on customers' levels and patterns of electricity usage. They reveal very limited information, however, about why or how customers respond to the rates or treatments—in particular, the effect of prices—to support drawing definitive and universally applicable conclusions about the effects of a single price structure compared to the behavioral influences of alternative rate structures.

Moreover, by definition, dynamic rate structures involve rates that change over time; the models discussed in Section 3 that measure load impacts are silent about how these changes in prices affect customers' levels and patterns of electricity usage. Measures of price response are a critical component in an overall program assessment. Price effects must be accounted for explicitly in order to make informed and reliable judgments about the impacts of dynamic pricing structures. Utilities routinely use some measure of price response in forecasting electricity demand for the purposes of setting rates and capacity planning, and marketing departments need this information to recruit customers to alternative rates.

To obtain estimates of price response, it is necessary to estimate electricity demand functions and derive from them price elasticities.[31] Price elasticities differ from raw measurements of changes in electricity consumption (load impacts) in that they measure the percentage change in electricity demand resulting from a 1% change in the price of electricity. In other words, the measured impact can be tested for significance and extended to treatment levels other than those tested.

We turn now to a discussion of the methods used to estimate these price effects. The models of price response that are an essential part of an assessment of dynamic rate structures provide estimates of two measures of the price elasticity of demand: an own-price elasticity demand for electricity and an elasticity of substitution between two time periods of the day.

[31] Throughout this discussion, it is assumed that the reader has some background in economics and demand modeling. For completeness and to assist those with less background in economics, much of the theory underlying the methods and models is provided in footnotes and technical appendices.

The first measure is an own-price elasticity of demand for electricity, or the percentage change in electricity usage during some period of time (for example, day, week, billing period, or season) that results from a 1% change in the price of electricity during that same period of time. The own-price elasticity of demand is characterized by some analysts as a measure of an overall effect on usage due to customer's movement from a fixed rate to a dynamic rate structure. The effect may be an increase or a decrease in usage, depending on the pricing structure and how the customer values electricity. Because the demand for electricity is expected to rise as the price of electricity falls, the own-price elasticity of demand should be negative.

The elasticity of substitution quantifies load shifting between time periods within a day. To measure load shifting, most recent studies have adopted a convention of dividing the day into a peak and an off-peak period.[32] Formally, to characterize the way in which changes in prices cause a change in usage across these periods, we estimate an elasticity of substitution between the peak and off-peak period—which is defined as the percentage change in ratio of the peak to off-peak electricity usage resulting from a 1% change in the ratio of off-peak to peak electricity prices.[33] By defining the elasticity of substitution in this way (for example, the ratio of off-peak to peak price of electricity), positive values of the elasticity of substitution are consistent with the underlying theory of demand. As the ratio of the off-peak price falls relative to the peak price, this inverse price ratio falls, and the ratio of peak to off-peak usage would be expected to fall as well.

The following paragraphs describe the estimation of these two types of price elasticities with several examples that involve more than one treatment. In many applications, separate models could be specified to estimate the elasticities for each treatment.[34] We believe, however, that there are advantages to estimating these elasticities by pooling the data and estimating a single model when possible. This modeling approach facilitates tests of hypotheses concerning the differences in the elasticities across treatments by first selecting a "reference treatment group" and then capturing the differential treatment effects by introducing dummy (or zero-one categorical) variables for the remaining treatment groups.[35] In situations in which there are fewer

---

[32] This *peak period* is most often defined as a period of hours in which prices in a dynamic rate, such as an RTP rate, are generally high. In the case of rates with an overlay option, such as a CPP, the *peak period* is defined as those hours for the CPP event. Defined in this way, this peak period may or may not be coincident with the system peak demand.

[33] Because there may be statistical correlations between the equation used to estimate the daily own-price elasticity of demand and the equation needed to estimate the elasticity of substitution, the two equations can be estimated jointly using a seemingly unrelated regressions (SUR) technique to obtain the most efficient parameter estimators.

[34] Assuming that one has estimates of identically specified regression models for each treatment group, one can test for the hypothesis that the regression coefficients that capture the treatment effects differ for any pair of treatments. For large, independent samples—and under the null hypothesis that two regression coefficients are equal—the difference between the two coefficients divided by the square root of the sum of the two squared standard errors for the coefficients follows a standard unit normal distribution (Clogg et al.,1995, pp. 1273–1277).

[35] Gujarati (1970a, b) was among the first to discuss the use of dummy variables to test the equality of pairs, or larger sets of coefficients in two or more regressions, where these regressions could be for treatment groups of equal or unequal size. This approach is consistent with models estimated using customer fixed-effects estimators, and it is also possible to correct for heteroskedasticity within or across treatments using any of several well-known methods for doing so (Gujarati, 1970b). Furthermore, it is also well known that the separate regressions for each treatment group can be retrieved from a single regression model for all treatment groups combined by adding dummy variables for all but the reference group as well as the interaction terms between each of these dummy variables with

treatments or just a single treatment, the model specifications are, of course, simpler. Several additional examples of alternative specifications of these models used to estimate own-price elasticities and elasticities of substitution are provided in the pilot studies reviewed in EPRI reports 1023562 and 1024867.

4.1 Estimating Own-Price Elasticities of Demand

In many empirical analyses of dynamic rates, estimates of own-price elasticities of demand are derived from log-linear demand models in which the natural logarithm of average hourly or total daily usage is modeled as a function of the natural logarithm of the average of that day's hourly prices. This model is convenient because the coefficient on the logarithm of price can be interpreted directly as the own-price elasticity of demand.

In its simplest model specification, the own-price elasticity of demand is constant at all prices and usage levels. Despite the fact that a log-linear demand model is a rather *ad hoc* specification, it has nonetheless been used widely in the empirical estimation of demand equations for products ranging from agricultural and food commodities and other consumer goods to energy.[36] Because its application to dynamic rates is typically to estimate very short-run electricity demand models (for example, daily demand), prices of substitutes such as other forms of electricity are assumed not to vary. In such analyses, there is no reason to consider the demand for more than this single product.

Were it necessary, however, to apply this constant elasticity model to a complete demand system for more than two products, it would be integrable only if it were generated by an additive logarithmic utility function—in which case the own-price elasticities would be equal to minus unity, and all cross-price elasticities would be equal to zero (LaFrance 1986). LaFrance (1986) goes on to show that a constant elasticity of demand model is only slightly less restrictive for incomplete systems of demand equations. Despite these restrictions needed for such a model to conform to the conditions required for complete demand systems, the application of a log-linear demand model has stood the test of time—particularly as an approximation for forecasting or for use in estimating very short-run demand curves in which prices of substitute or complementary commodities are unlikely to change.

4.1.1 The Log-Linear Model

It is instructive to consider a specific example. Let us assume that we have a pooled data set that consists of a time series of cross-sections and that there are $t = 1,\ldots,n$ daily observations for each customer $i = 1,\ldots,m$. Let us also assume that there are three rate treatments in this pilot:

---

each of the regressors in the single equation (for example, Gujarati, 1970b and Clogg et al., 1995). Finally, one can still test for the equality of treatment effects across groups even if not all of the interaction terms between the dummy variables and the regressors are included in the model. In this case, however, the tests are conditioned on the assumption that effects of certain regressors do not differ across groups.

[36] It is *ad hoc* in the sense that it is not structurally consistent with utility theory, so the interpretation of the parameters as elasticities is representative—although not necessarily fully so—of electricity demand.

1. Customers in the first treatment face an RTP rate in which the 24-hourly prices are announced a day ahead.
2. Customers in the second rate treatment generally face the RTP prices, but there are a certain number of critical peak event days that can be called during the summer months. On RTP/CPP event days, Treatment 2 customers pay very high CPP prices during the peak hours of the day.
3. Finally, customers in Treatment 3 face RTP prices on all but the critical event days; during the event hours they are **paid** the high CPP prices to reduce load, called a *peak time rebate* (RTP-PTR). Because prices differ by hour for customers in each of the rate treatments, average daily prices will as well.

Therefore, we can specify a simple log-linear model to estimate the daily own-price elasticity of demand for electricity as:

Eq. 13: $lnK_{ti} = \alpha + \beta lnP_{ti} + u_{ti}$

where: K_t is the daily demand measured in kWh, and P_{ti} is a measure of the daily price of electricity in $/kWh (perhaps the average hourly price on a given day), and u_{ti} is an error term.

In this form, the model restricts the own-price elasticity of demand to be constant for every day and for every customer, regardless of differences in the weather or in customer or premise characteristics. As mentioned, this limitation can be addressed in the model specification by adding other variables to the model as interaction terms with the price variable.[37]

For example, to account for the effect of different weather conditions on the daily demand for electricity—and for its effect on price response—we can augment Equation 4-1 in several ways. The resulting specification is more complex but complete (for example, Equation 4-2). We begin by adding two weather variables. The first, CD_{ti}, measures cooling degrees, and it enters as an intercept shifter, controlling for differences in daily electricity demand as temperature changes. The second, H_{ti}, on the other hand, takes on the binary values of unity for hot days and zero otherwise, where a hot day is defined as one in which a heat index is above 85°F (29°C). In the specification of the demand equation, this variable is included as an interaction term that accounts for a change in the own-price elasticity of demand between hot and cool days.[38]

---

[37] In this model, one could also include dummy variables to control for the effects of groups of customers or customer-specific characteristics. These demand models can include customer fixed effects in a manner similar to that used in the models of load impacts. Therefore, if we also exploit the pooled nature of the data and estimate a fixed-effects model, the only way to identify the effects of explanatory variables that are time-invariant for each customer is to assume that these customer or premise characteristics (for example, AC) affect usage primarily through their interaction with treatments and/or factors (for example, weather) that vary over time. Therefore, these effects can be captured by the coefficient associated with variables created by constructing interaction terms between these types of variables with those variables that represent the various treatments. Although we do not specify this model specifically as a fixed-effects model, the model specification is consistent with this way of capturing the effects of these customer-specific characteristics. If we were to specify a fixed-effects model, we would add a term μ_i to the model in Equation 4-1 and use a fixed-effects estimator to estimate the model.

[38] The specific construction of these two weather variables is primarily for purposes of illustration and follows Boisvert et al. (2007).

Next, assume that the binary variable (AC_i) takes on a value of unity if the customer has central AC, and zero otherwise. Assume further that we suspect that the daily demand for electricity will depend on the rate treatment. We can choose the RTP customers as the reference group and include (0, 1) dummy variables for customers in the other rate treatments to test the hypothesis that own-price elasticities will differ by rate treatment. Therefore, we specify a (0, 1) variable for customers on an RTP-CPP rate (CPP_i) and another (0, 1) variable (PTR_i) for customers on an RTP-PTR rate. Finally, we specify a (0, 1) variable (Low_i). If the customer is a low-income customer, this variable is unity; otherwise, it is zero.

By including these additional variables in the original model in Eq. 13, the estimating equation would now take the following form:

Eq. 14: $lnK_{ti} = \alpha + cc(CD_t) + [\beta + ac(AC_i) + hot(H_t) + cp(CPP_i) + pt(PTR_i) + ww(Low_i)]lnP_{ti} + u_{ti}$

To estimate this model, there are now eight variables in the model—one dependent variable, lnK_{ti}, and seven independent variables: (CD_t), lnP_{ti}, (AC_i)lnP_{ti}, (H_t)lnP_{ti}, (CPP_i)lnP_{ti}, (PTR_i)lnP_{ti}, and (Low_i)lnP_{ti}.[39]

The parameters associated with these variables are evident from Eq. 14. Because the variable (CD_t) measures the cooling degrees, one might expect that as this variable increases *ceteris paribus*, there would be an increase in the demand for electricity at any price. We would therefore expect the estimated value of the coefficient on this variable to be positive, $cc > 0$, shifting the demand curve upward.

To identify the own-price elasticity of demand for electricity, recall that it should be negative. We would expect that the coefficient on the logarithm of price, β, would be negative. We can identify the effects of the other variables on the daily own-price elasticity of demand as follows. Define the effective own-price elasticity of demand as ε^E. Accordingly, we have:

Eq. 15: $\hat{\varepsilon}^E = \hat{\beta} + \widehat{ac}(AC_i) + \widehat{hot}(H_t) + \widehat{cp}(RTP_i) + \widehat{ptr}(PTR_i) + \widehat{ww}(Low_i)$

where the superscript "hat" (^) indicates an estimated coefficient. We might expect customers who rely on AC to keep cool on hot summer days to be less willing to reduce load in response to a price increase than customers with no AC. Therefore, the own-price elasticity of demand would be lower (for example, a negative number closer to zero), implying that $ac > 0$.
Furthermore, because customers on both RTP and PTR rates are faced with different, and substantially higher, prices during some hours of the day, one might expect these customers to be more likely to change daily demand in response to prices—in part because they may be more aware of daily changes in prices and so on. Therefore, we might expect the coefficients on these two variables to be negative, $cp < 0$ and $ptr < 0$, increasing the absolute value of the own-price elasticity of demand.

---

[39] Regardless of how these models are estimated (for example, with or without the fixed-effects estimator), standard errors should be estimated using the "cluster" option, available in most statistical packages, to account for correlation in errors at the customer level over time.

If one expects that the customer's willingness to reduce demand as the price rises is lower on hot days, one would expect the coefficient on the variable for H_t to be positive, *hot* > 0.

Finally, it is a bit more difficult to speculate on the sign of the coefficient on the variable representing low-income customers. On one hand, they may be more likely to reduce usage when prices rise in order to reduce their bill. However, they may also be less able to do so, relative to high-income customers, because they may have an older and less energy-efficient stock of appliances or only a window AC unit. For these reasons, the coefficient, *ww*, could be positive or negative.

Based on this discussion, it is clear that these additional factors can work to increase or decrease the own-price elasticity of demand. The net effect on this measure of price response will depend on the relative sizes of the positive and negative coefficients.

In this model, there are now several separate own-price elasticities of demand (one for each combination of values for the six dummy variables included as slope shifters in Eq. 13). As a base of comparison, assume that the customer is not low income, is on neither an RTP-CPP nor an RTP-PTR rate, and does not have AC. If it is not a hot day, then $AC_i = 0$; $H_t = 0$; $CPP_i = 0$; $PTR_i = 0$; and $Low_i = 0$. Because all of the dummy variables are zero, the elasticity of demand is simply given by:

Eq. 16: $\hat{\varepsilon}^E = \hat{\beta}$

Any of the combinatorial estimates of the elasticity of demand can be depicted in a similar fashion. The coefficients that appear in a particular estimate depend on which of the dummy variables take on values of unity and which take on values of zero. For illustration, three of the possible elasticities of demand are listed next.

If the customer is not low income but is an RTP-CPP customer with AC, and it is a hot day, then $AC_{ti} = 1$; $H_{ti} = 1$; $CPP_{ti} = 1$; $PTR_{ti} = 0$; and $Low_{ti} = 0$. Therefore:

Eq. 17: $\hat{\varepsilon}^E = \hat{\beta} + \widehat{ac} + \widehat{cp} + \widehat{hot}$

If the customer is not low income, is on neither an RTP-CPP nor an RTP-PTR rate and does not have AC, and it is a hot day, then $AC_{ti} = 0$; $H_{ti} = 1$; $RTP_{ti} = 0$; $CPP_{ti} = 0$; $Low_{ti} = 0$; and $EV_t = 0$. Therefore:

Eq. 18: $\hat{\varepsilon}^E = \hat{\beta} + \widehat{hot}$

If it is a hot day, but it is an event day, and the customer has AC, is low income, and is on neither an RTP nor a CPP rate, then $AC_{ti} = 1$; $H_{ti} = 1$; $CPP_{ti} = 0$; $PTR_{ti} = 0$; $Low_{ti} = 1$; and $EV_{ti} = 0$. Therefore:

Eq. 19: $\hat{\varepsilon}^E = \hat{\beta} + \widehat{ac} + \widehat{hot} + \widehat{ww}$

In each of Eq. 17, 18, and 19, the net result on the customer's willingness to reduce load in response to price relative to the base or reference case of Eq. 16, as measured by the elasticity of demand, depends on the extent to which the sum of the relevant positive effects (for example, $ac > 0$ and/or $hot > 0$) is partially or totally offset by negative effects (for example, $cp < 0$ and $ptr < 0$), remembering that the sign on the low-income variable, *ww*, could be positive or negative.

In this example of a daily demand function for electricity, we have focused on only one intercept shifter and six slope shifters formed by interaction terms with the logarithm of price. There are, of course, several other variables that could logically be included in the demand equation, depending on the types of rate and information treatments in the pilot and the type of socioeconomic data collected in any customer survey. For example, it may be logical to include slope shifters for customers who are given access to enabling technology and/or educational information about certain dynamic rates.

4.1.2 The Semi-Log Model

In addition to the log-linear model of demand, it is not uncommon for analysts to also estimate semi-log demand models in which the natural logarithm of usage is regressed on the price of electricity—not on the logarithm of the electricity price. This model is distinguished from the log-linear model in that the own-price elasticity of demand is a function of the price level (for example, the estimated coefficient on the price term multiplied by the price), indicating that demand is more elastic at high prices.

These models are also typically specified to include some conditioning variables such as intercept shifters and/or as interaction terms with the logarithm of price. By including these additional variables, the demand curve is shifted and/or twisted. Therefore, the own-price elasticities in this model will also differ depending on the conditioning variables—such as weather, customer or premise circumstances, whether the customer has access to enabling technology, and whether it is an event day.[40]

The semi-log demand model consistent with the log-linear demand model in this example is specified below; all variables are defined in the log-linear model above. The model, excluding any shifter variables, is given by:[41]

Eq. 20: $lnK_{ti} = \alpha + \beta P_{ti} + u_{ti}$

Once the additional variables are added to form the respective intercept and slope shifters, the model becomes:[42]

---

[40] Examples of such specifications are found in EPRI's review of recent dynamic rate pilots (EPRI 1023562).

[41] As discussed along with the log-linear model, we can also exploit the pooled nature of the data. If we were to do so by specifying a fixed-effects model, we would add a term μ_i to the model in Equation 4-9 and use a fixed-effects estimator to estimate the model.

[42] Again, regardless of how these models are estimated (for example, with or without the fixed-effects estimator), standard errors should be estimated using the "cluster" option, available in most statistical packages, to account for correlation in errors at the customer level over time.

Eq. 21: $\mathrm{lnK_{ti}} = \alpha + \mathrm{cc(CD_t)} + [\beta + \mathrm{ac(AC_i)} + \mathrm{hot(H_t)} + \mathrm{cp(CPP_i)} + \mathrm{ptr(PTR_i)} + \mathrm{ww(Low_{ti})]P_{ti}} + \mathrm{u_{ti}}$$lnK_{ti} = \alpha + cc(CD_t) + [\beta + ac(AC_i) + hot(H_t) + cp(CPP_i) + ptr(PTR_i) + ww(Low_{ti})]P_{ti} + u_{ti}$

The equation for the own-price elasticity of demand is now as shown in Equation 4-10, where the "hat" (^) represents an estimated coefficient:

Eq. 22: $\hat{\varepsilon}^E = [\hat{\beta} + \widehat{ac}(AC_i) + \widehat{hot}(H_t) + \widehat{cp}(CPP_i) + \widehat{ptr}(PTR_i) + \widehat{ww}(Low_i)]P_{ti}$

Eq. 22 differs from Eq. 15 in that all terms are multiplied by price (P_{ti}). Therefore, the own-price elasticity of demand for electricity varies directly with the price of electricity; in the log-linear model, the own-price elasticity does not depend only on the price but is determined by the price structure employed. In this model, there are again several separate own-price elasticities of demand—one for each combination of (0, 1) values for the six dummy variables. They are formed similarly to Eq. 16 through Eq. 19.

The interpretation of the results would be similar as well. For simplicity, we have used the same symbols to represent the coefficients for the variables in the model. However, if both models in Eq. 14 and Eq. 21 were estimated from the same data, the estimated parameters would certainly differ because the price variable in Eq. 14 is lnP_t, whereas the price variable is simply P_t in Eq. 21. These estimates of the own-price elasticities would differ both because of the different parameter estimates and the fact that the price also appears in Eq. 22 but not in Eq. 16.

4.1.3 Comparison to Load Impact Analysis

In recent dynamic rate pilots, the empirical specifications of these demand models (whether log-linear or semi-log) employed to estimate own-price elasticities of demand are strikingly similar to the specification of models used to estimate load impacts. They share many aspects of their mathematical structure and employ the same estimation techniques. The primary distinction is the inclusion of prices as explanatory variables in the demand specifications; for this reason, some analysts argue that these demand models also provide a better measure of a conservation effect of dynamic rates (for example, how the overall or average daily price changes associated with a change in relative prices affects total daily energy consumption).

Because of the lack of substitutes for electricity in the short term, these models normally do not include the prices for substitute forms of energy. We might also expect estimates of this notion of the own-price elasticity of demand to be rather small in absolute value (perhaps relatively close to zero). For the most part, the empirical evidence in recent pilot studies, as in past studies, confirms this hypothesis.[43]

Despite this empirical evidence, analysts are often surprised that the measured conservation effects (for example, the reduction in daily usage) from these new rates are extremely small or that, in some cases, daily demand increases. These results may in fact be consistent with the new rate structure, which, in the utility's attempt to retain the revenue neutrality of the rate, would

[43] Again, see EPRI's review of recent dynamic rate pilots (EPRI reports 1023562 and 1024867).

lead to an actual reduction in daily average hourly price of electricity, albeit it a rather small one.[44]

4.2 Measures of Load Shifting: Estimating Elasticities of Substitution

The elasticity of substitution measures the **percentage change** in the ratio of electricity usage during these peak hours to the electricity usage in these non-peak hours resulting from a **1% change** in the ratio of the off-peak to peak prices of electricity. (The bold emphasizes the different mathematical structure of how price response is portrayed.)

To estimate this measure of load shifting between time periods within a day, one must define the peak and off-peak electricity by specifying what hours of the day are assigned to each in addition to other sorting criteria, such as whether those periods are applicable to weekends or apply year-round. In most pilots in which elasticities of substitution are estimated, the peak periods are defined for purposes of estimation to coincide with the event hours for the CPP rates.[45]

To estimate the elasticity of substitution, one can assume that the underlying model of demand (derived from a consumer utility function) takes on one of two basic algebraic forms: a standard constant elasticity of substitution (CES) model or a more flexible form, the generalized Leontief (GL) model. Both the GL and CES models accommodate tests of the null hypothesis that the elasticity of substitution is zero. The capacity to test the elasticity of substitution against the hypothesis that its value is zero is a particularly attractive feature because some segments of residential customers or certain types of commercial and industrial customers are unable or unwilling to change electricity usage from one part of the day to another.

The GL model is more flexible in that the elasticity of substitution between peak and off-peak electricity can differ among customers as well as daily. This characteristic of the model is a distinct advantage if there is good reason to believe 1) that the inherent ability or capacity of some customers to shift load is substantially different from that of others and 2) that these shifts may be in response to nominal peak prices as well as off-peak to peak price ratios. These differences can be masked in CES-type models because they are washed out by averaging out individual customer's measures by enforcing the assumptiom of a constant measure of the elasticities of substitution.[46] The derivations of the elasticities of substitution for each of these demand models are given in Appendix D.

---

[44] Revenue neutrality is achieved when, based on the class average load profile, the conventional rate and the dynamic rate yield the same bill for that average level of load. That is, unless customers on the average alter their usage on the dynamic rate, they will pay the same. However, this revenue neutrality is premised on the class load. Individual customers have a greater or lesser degree of equality in bills depending on their particular load profile relative to that of the class average. Accordingly, some may realize an opportunistic gain simply by subscribing—their bill is lower without responding.

[45] These peak hours may or may not be coincident with the system peak demand. In cases in which the CPP customers face an RTP rate on non-event days, the RTP prices are generally higher in these hours than during other hours of the day.

[46] In a study of large commercial and industrial customers, Boisvert et al. (2007) found this to be the case. The extent to which this may also be true for residential customers is an open question, although in the evaluation of the

4.2.1 The CES Model

As its name implies, the elasticity of substitution derived from the CES model is constant, and one can estimate the elasticity of substitution by creating a data set of daily observations for all customers, and regressing the ratio of peak to off-peak electricity usage on the ratio of off-peak to peak prices.[47] Through the inclusion of interaction terms between the ratio of peak to off-peak prices and other independent variables, the model specification can be enriched to allow for the elasticity of substitution to differ by such things as customer or premise characteristics, weather, whether a customer is a CPP or PTR participant, and whether the customer has enabling technology .[48]

To illustrate, we discuss a CES model for the elasticity of substitution between peak and off-peak electricity usage that is consistent with variables used to generate intercept and slope shifters in the example daily demand Equations 4-2 and 4-9. To reiterate, let us assume that we have a pooled data set that consists of a time series of cross-sections and that there are $t = 1,\ldots,n$ daily observations for each customer $i = 1,\ldots,m$. Customers in the first of three rate treatments face an RTP rate in which the 24-hourly prices are announced a day ahead. Customers in the second rate treatment generally face the RTP prices, but there are a certain number of critical peak event days that can be called during the summer months. On these event days, Treatment 2 customers pay very high CPP prices during the peak hours of the day (RTP-CPP). Finally, customers in Treatment 3 face RTP prices on all but the critical event days, but during the event hours they are **paid** the high CPP prices to reduce load—called a *peak time rebate* (RTP-PTR). Then, we begin with the model (as derived in Appendix D):[49]

---

recent ComEd CAP program, these differences were not large enough to warrant the additional computational complexity of the GL model (EPRI report 1023644).

[47] This within-day elasticity of substitution between peak and off-peak electricity can also be estimated using a nested CES (NCES) model. This model is unique in that it allocates a customer's electricity expenditures separately within a day and between days. The within-day elasticity of substitution is constant, but it can differ from the constant elasticity of substitution between days. In the NCES structure, the **between-day** elasticity of substitution is important if its value differs from that of the **within-day** elasticity of substitution. If the between-day elasticity is larger than the corresponding within-day elasticity on a CPP event day, for example, load is shifted away from event days. This would be a questionable result for residential customers whose load during event days is likely to be largely discretionary. However, for some commercial or industrial customers who shift load to non-event days to make up for lost production, this model may well be appropriate. One early application of this NCES model to large industrial customers is by Herriges et al. (1993).

[48] By allowing the elasticity of substitution to differ based on these characteristics, we effectively account for differences in the curvature of the indifference curves of the customer's underlying utility function. As the degree of the curvature increases, the elasticities of substitution fall (substitution possibilities are fewer); if they become less curved, the elasticities of substitution increase.

[49] In this model, one could also include dummy variables to control for the effects of groups of customers or customer-specific characteristics. However, if we also exploit the pooled nature of the data and estimate a fixed-effects model, the only way to identify the effects of explanatory variables that are time-invariant for each customer is to assume that these customer or premise characteristics (for example, AC) affect usage primarily through their interaction with treatments and/or factors (for example, weather) that vary over time. Therefore, these effects can be captured by the coefficient associated with variables created by constructing interaction terms between these types of variables with those that represent the various treatments. Although we do not specify this specifically as a fixed-effects model, the model specification is consistent with this way of capturing the effects of these customer-specific characteristics. If we were to specify a fixed-effects model, we would add a term μ_i to the models in Equation 4-11 and use a fixed-effects estimator to estimate the model.

Eq. 23: $\ln \frac{(K_{pti})}{(K_{oti})} = \gamma + \sigma \ln \frac{(r_{oti})}{(r_{pti})} + \epsilon_{ti}$

where K_{pti} and K_{opi} are electricity usages during the peak and off-peak periods, respectively, and r_{ti} and r_{oti} are electricity prices during the peak and off-peak periods, respectively. In this form, the model restricts the elasticity of substitution to be constant for every day and every customer in all three treatments, and it is equal to σ.

It is also important to note that the price term is expressed as the inverse price ratio. This explains why one would expect elasticities of substitution to be positive (for example, $\sigma > 0$).[50] That is, in moving from a flat rate to an RTP rate or a TOU rate, the peak price would normally be above the flat rate and the off-peak price would normally be below the flat rate. These changes lead to a decrease in the off-peak to peak price ratio, which in turn leads to a substitution of off-peak load for peak load—resulting in a reduction in the ratio of peak to off-peak load.

We can account for the effect of different weather conditions on the demand for peak and off-peak electricity by augmenting Eq. 23 with two weather variables. The first, variable (CD_t), measures cooling degrees, and it enters as an intercept shifter, controlling for differences in peak to off-peak usage as temperature changes. Variable H_t takes on the binary values of unity for hot days and zero otherwise, where a hot day is defined in some measurable terms, for example, when a heat index is above 85°F (29°C).

Furthermore, assume that the binary variable (AC_i) takes on a value of unity if the customer has central AC and is zero otherwise. We specify a (0, 1) variable (Low_{ti}) if the customer is a low-income customer, in which case this variable is assigned a value of unity; otherwise it is zero. Finally, in this model to estimate elasticities of substitution, it is particularly important to test for differences in the effect of various rate treatments on the willingness of customers to shift load from peak to off-peak periods. Therefore, we also specify a (0, 1) variable for customers on an RTP-CPP rate (CPP_i) and another (0, 1) variable (PTR_i) for customers on an RTP-PTR rate. By adding these variables to the original model in Eq. 23, the model to be estimated is now as shown in Eq. 24:

Eq. 24:

$$\ln \frac{(K_{pti})}{(K_{oti})} = \gamma + cd(CD_t) + [\sigma + a(AC_{ti}) + h(H_t) + \mathrm{c}(CPP_i) + c(PTR_i) + w(Low_i)]\ln \frac{(r_{oti})}{(r_{pti})} + \epsilon_{ti}$$

To estimate this model, there are now eight variables: one dependent variable and seven independent variables along with a constant term γ.

The dependent variable is $\ln \frac{(K_{pti})}{(K_{oti})}$. The independent variables are as follows:[51]

[50] Some analysts do not invert the price ratio on the right-hand side; this alternative convention results in the same estimated numerical value but with a negative sign.

[51] Again, regardless of how these models are estimated (for example, with or without the fixed-effects estimator), standard errors should be estimated using the "cluster" option, available in most statistical packages, to account for correlation in errors at the customer level over time.

(CD_t); $\ln\frac{(r_{oti})}{(r_{pti})}$; $(AC_i)\ln\frac{(r_{oti})}{(r_{pti})}$; $(H_t)\ln\frac{(r_{oti})}{(r_{pti})}$; $(CPP_i)\ln\frac{(r_{oti})}{(r_{pti})}$;
$(PTR_i)\ln\frac{(r_{oti})}{(r_{pti})}$; $and\ (Low_i)ln\frac{(r_{oti})}{(r_{pti})}$

The parameters associated with these variables are evident from Eq. 24. Because the variable (CD_{ti}) measures the cooling degrees, one might expect that as this variable increases, *ceteris paribus*, there would be an increase in the ratio of the hourly peak to off-peak electricity use and the value of the coefficient on this variable is therefore positive, $cd > 0$. To interpret the other variables, it is essential to recognize that they affect the elasticity of substitution in the following way. Define the effective elasticity of substitution as σ^E. We have:

Eq. 25: $\hat{\sigma}^E = \hat{\sigma} + \hat{a}(AC_{ti}) + \hat{h}(H_{ti}) + \hat{c}(CPP_{ti}) + \widehat{pt}(PTR_{ti}) + \hat{w}(Low_{ti}))$

where the "hat" (^) indicates an estimated coefficient.

We might expect those customers who rely on AC to keep cool on hot summer days to be less willing to shift load off-peak than customers with no AC. The elasticity of substitution would be lower, and $a < 0$. If one expects that the customer's willingness to shift load from peak to off-peak falls on hot days, one would expect the coefficient on the variable for (H_t) to be negative as well: $h < 0$.

Because customers on both RTP and CPP rates are faced with higher, and variable, prices during peak hours of the day, one might expect these customers to be more willing to shift load from peak to off-peak hours. We might expect the coefficients on these two variables to be positive, $c > 0$ and $pt > 0$, increasing the elasticity of substitution. As with the daily demand model, it is not easy to speculate on the "sign" of the coefficient on the variable (*Low*). In efforts by low-income customers to reduce their electricity bill, one might expect them to be more willing to shift load from peak to off-peak periods in response to higher peak prices, but their ability to do so may be hampered by the nature of their stocks of appliances. The net effect could be determined only empirically.

In this model, there are now several separate elasticities of substitution (one for each combination of values for the five dummy variables). As previously, we assume the following as a base of comparison: 1) the customer does not have AC, 2) it is not a hot day or an event day, 3) the customer is not low-income, 4) the customer is not on an RTP-CPP rate, and 5) the customer is not on an RTP-PTR rate. Therefore, $AC_i = 0$; $H_t = 0$; $CPP_i = 0$, $PTR_i = 0$; and $Low_i = 0$. Because all of the dummy variables are zero, the elasticity of substitution is simply given by:

Eq. 26: $\hat{\sigma}^E = \hat{\sigma}$

Any of the other number of separate estimates of the elasticity of substitution can be depicted in a similar fashion. The coefficients that appear in a particular estimate depend on which of the dummy variables take on values of unity and which take on values of zero. The net result on the willingness to shift load relative to this "base" customer, as defined in Equation 4-14, depends on the extent to which the sum of the relevant negative effects (for example, $a < 0$ and $h < 0$) are partially or totally offset by positive effects (for example, $c > 0$ and $pt > 0 > 0$), recognizing that the parameter w may be positive or negative. For illustration, three of the possible elasticities of substitution are listed next.

If the customer has AC and is on the RTP-CPP rate, and it is a hot day, then $AC_i = 1$; $H_t = 1$; $CPP_i = 1$; $PTR_i = 0$; $Low_i = 0$; and $EV_t = 1$. Therefore:

Eq. 27: $\hat{\sigma}^E = \hat{\sigma} + \hat{a} + \hat{h} + \hat{c}$

If the customer is not low-income, is neither on an RTP-CPP rate nor on an RTP-PTR rate, does not have AC, and it is a hot day, then $AC_i = 0$; $H_t = 1$; $CPP_i = 0$, $PTR_i = 0$; and $Low_i = 0 = 0$. Therefore:

Eq. 28: $\hat{\sigma}^E = \hat{\sigma} + \hat{h}$

If it is not a hot day, and the customer is on an RTP-CPP rate, is a low-income customer, and has AC, then $AC_i = 1$; $H_t = 0$; $CPP_i = 1$, $PTR_i = 0$; and $Low_i = 1$. Therefore:

Eq. 29: $\hat{\sigma}^E = \hat{\sigma} + \hat{a} + \hat{c} + \hat{w}$[52]

4.2.2 The GL Model

In contrast to the CES model, the GL model is more flexible in that the elasticity of substitution between peak and off-peak electricity can differ by customer and by day. This characteristic of the model is a distinct advantage if, as emphasized previously, there is reason to believe that the inherent ability or capacity of some customers to shift load is substantially different from that of other customers and differs among days—the differences are therefore masked in CES-type models because they end up being "averaged" out in the constant measures of the elasticities of substitution.

With the added flexibility of the GL model, it is also possible to test whether the elasticity of substitution value is zero and also to test if the elasticities of substitution differ by customer and by day. The procedures for estimating these elasticities based on the GL model are derived in detail in Appendix D.

Flexibility comes at a substantial cost in terms of the computational burden. Because the elasticities of substitution can differ by customer and by day, separate equations must be estimated for each customer. For this reason, some information is lost by not exploiting the panel

[52] In this example of the CES specification of the elasticity of substitution, it is also important to emphasize that, in these equations to estimate elasticities of substitution of peak for off-peak electricity, we have again focused on only one intercept shifter and five slope shifters formed by interaction terms with the logarithm of price. There are, of course, several other variables that could be logically included in the demand equation, depending on the types of rate and information treatments in the pilot and the type of socioeconomic data collected in any customer survey. For example, it may be logical to include slope shifters for customers who are given access to enabling technology and/or educational information about certain dynamic rates.

nature of the data. Furthermore, the process begins with the estimation of the ratio of daily peak to off-peak expenditure shares using a method of nonlinear least squares. From the results of these estimated equations, it is possible to calculate predicted shares and use them in the calculation of the elasticities of substitution. As is shown in Appendix D, it is quite easy to incorporate weather variables into equations to estimate the expenditure shares; in so doing, the elasticities of substitution are also functions of the weather variables.

The GL procedure yields a distribution of daily elasticities of substitution for each customer. However, this procedure does not identify any factors other than weather (such as customer and premise characteristics, the presence of enabling technology, or education) that explain why these elasticities of substitution for each customer differ by day. This can be accomplished by estimating another regression model. This second regression is specified by first creating a pooled data set that includes the daily estimates of the elasticities of substitution for all customers. This data set is then merged with customer-level data for customer and premise characteristics, the presence of enabling technology, education, and so on. The estimates of the daily elasticities of substitution are then regressed on these various exogenous factors related to, for example, customer and premise characteristics.

These procedures for constructing and estimating an additional regression model are also developed in detail in Appendix D under the heading of a “meta-analysis.” It is clear that one must be convinced that the elasticities of substitution differ substantially across customer and time to justify the extra computational burden needed to use estimate elasticities of substitution based on the GL model. Such a justification may be more likely in the analysis of large, diverse commercial and industrial customers than in an application involving residential customers.

5. Summary and Conclusions

Price and demand response—particularly the latter—have experienced resurgence in the past decade. Widespread utility interest over the past decade to invest in AMI have compelled regulators and other policymakers to assess the validity of purported non-operational benefits from price and/or demand response enabled by such investments, in part because many prior research efforts had not explicitly focused on such types of opportunities in a systematic fashion. The new wave of pilots and experiments is designed to quantify the effects of a variety of behavioral inducement programs on participants' electricity consumption patterns.
In implementing these experiments and pilots, utilities used a variety of design, execution, and evaluation methods. These differences have made it difficult to compare readily and usefully the conclusions drawn across studies, leaving many important aspects about customers' use and value of electricity unanswered. In response, many policymakers continue the cycle by compelling their regulated utilities to undertake their own unique research efforts to characterize the way in which behavioral inducements influence electricity consumption (EPRI report 1025856, 2012).

The potential exists for these initiatives to fill many of these remaining gaps in our knowledge and advance our understanding of electricity customer behavior. However, a substantial effort at this juncture to provide guidance on the research design, analysis and coordination across studies may eliminate any serious methodological shortcomings and avoid the squandering of scarce resources which would have resulted in a duplication of effort, missed opportunities, and/or misleading findings.

This report strives to define analysis protocols and practices for measuring the effects of time-based rates, enabling technology, and various other treatments on customers' levels and patterns of electricity usage. The exposition segments programs and their behavioral inducements in relation to whether they were event-based or non-event-based to illustrate the differences in methodological approaches to designing, implementing, and evaluating pilots that included these types of treatments. It focuses on experimental designs that will, in theory, produce the most robust estimates of impacts on demand reductions, demand shifting, and/or energy conservation: randomized control trials or randomized encouragement designs. However, it also identifies the conditions under which alternative impact estimation techniques could be applied. Several different models of price response are reviewed as ways to estimate accurately own-price elasticities and/or substitution elasticities, important metrics for extrapolating results to price levels not directly integrated into the pilot or experiment.

In addition to using the proper analysis methods, advancing our understanding of how price and other inducements alter electricity demand requires that analysts fully report the methods they used and the result they produced. This should include, at a minimum, a description of the study design, a description of the data, complete details of the empirical specification and econometric estimation, and the results derived from the estimated models. Documentation should include the following, at a minimum:

- Description of the sample frame and target population
- Description of all of the treatments applied
- Randomization or other assignment methods used
- Recruitment approach used
- Number of customers in each step of the enrollment and retention process (that is, the number solicited, the number recruited, the number that dropped out initially, the number that installed the required technology, and so on)
- Description of the actual implementation

Because all of these pilots and experiments rely on analysis techniques to arrive at an impact estimate, a description of the data employed in those analysis techniques is likewise critical for an outside reviewer to assess the robustness of the results. Specifically, the data description should include, at a minimum, the frequency and duration of the data used in the analysis, the group of customers for which the data were collected, and any other data (such as demographic data) that were collected.

The data description should also describe any data that are missing, imperfect, interpolated, or atypical in any way. Customers, dates, and times that were included or excluded should be described. Each empirical model should be specified in detail, including precise definitions of the reference load, the dependent and independent variables, and the interpretation of any coefficients that are to be estimated. At a minimum, these models should include an estimate of each rate, technology, and information treatment.

Documenting the statistical procedures used to estimate each model is another vital step in garnering credibility for the final set of reported impact estimates. In addition to a general model specification, any special treatment of the error structure and/or method of estimating standard errors of the estimated coefficients should be reported and a short statement justifying the use of these procedures provided.

The results from the estimation of each model should be detailed along with the standard errors on the coefficients and the results of any hypothesis tests. We recommend that each estimate be reported in the following four standardized formats (where the *impact hour* is the hour of interest for the evaluation, for example, event hours for CPP rates or daily peak hours for TOU rates) for ease of comparability across pilots:

- The average kWh reduction per customer per impact hour (for example, 0.5 kWh per weekday peak hour per customer)
- The average percent energy reduction per impact hour (for example, 40% per weekday peak hour per customer)
- The total energy conservation: the average kWh reduction per customer per month over all hours in that month (for example, $0.40 kWh per month per customer)
- The average percent energy reduction per month over all hours in that month (for example, 4% per month per customer)

In addition to aggregate load impact estimates, a variety of additional analyses associated with these pilots or experiments could be undertaken that would help the industry to better understand the expected benefits that could be derived from a full implementation of the elements included in the pilot or experiment.

For example, deriving load impact and/or elasticity estimates for different time windows (for example, for certain months or certain hours) or durations (for example, first year vs. second year) is important for understanding the degree of persistence in a customer's response. Segmenting the load impact and/or elasticity estimates for certain subsets of customers (for example, for high-energy usage customers or for low-income customers) helps utility program designers and marketing staff to effectively target the program to customers most likely to enroll and produce the largest load impacts.

Analyzing customer choice to join the behavioral inducement program and to remain with that program over time can help a utility to set reasonable expectations for recruitment and retention efforts, which—when taken in conjunction with load impact estimates—can help the utility to understand the total load impacts such a program can produce over time. Although not discussed here, these types of analyses are vital for the industry to gain a better and more detailed understanding of how different subsets of customers make choices to pursue such rates and programs and how they respond once exposed to them.

Analysts are charged with conducting the best possible scientifically directed evaluation and providing interpretations of the results. An experiment or pilot that is fully evaluated using the proper methods is a success, regardless of the outcome.

6. References

Alcott, H., Social Norms and Energy Conservation," *Journal of Public Economics* (95) 1082–1095, 2011.

Allen, R. G. D., *Mathematical Analysis for Economists*. St. Martin's Press, Inc., New York, NY: 1938 (reprinted in 1964).

Allison, P., *Logistic Regression: Using the SAS System*. The SAS Institute, Inc., Cary, NC: 1999.

Allison, P., *Fixed-Effects Regression Methods in SAS®*. The SAS Institute, Inc., Cary, NC: 2005.

Altman, D., "Comparability of Randomised Groups," *The Journal of the Royal Statistical Society, Series D (The Statistician*) (34): 125–136, 1985.

Anderson, D., D. Sweeney, and T. Williams, *Statistics for Business and Economics*. South-Western College Publishing, Cincinnati, OH: 1999.

Angrist, J., G. Imbens, and D. Rubin, "Identification of causal effects using instrumental variables," *Journal of the American Statistical Association* (81): 444–455, 1996.

Banerjee, A., E. Duflo, R. Glennerster, and D. Kothari, "Improving Immunization Coverage in Rural India: A Clustered Randomized Controlled Evaluation of Immunization Campaigns with and without Incentives," *British Medical Journal* (340):c2220, 2010.

Berndt, E., *The Practice of Econometrics: Classic and Contemporary*. Addison-Wesley Publishing Company, Reading, MA: 1991.

Boisvert, R. N., "The Translog Production Function: Its Properties, Its Several Interpretations and Estimation Problems," Department of Agricultural Economics, A.E. Res. 82–28, Cornell University, September 1982.

Boisvert, R., P. Cappers, C. Goldman, B. Neenan, and N. Hopper, "Customer Response to RTP in Competitive Markets: A Study of Niagara Mohawk's Standard Offer Tariff," *The Energy Journal* (28): 53–73, 2007.

Boisvert, R., B. Neenan, and J. Robinson, *Residential Electricity Use Feedback: A Research Synthesis and Economic Framework*. EPRI, Palo Alto, CA: 2009. 1016844.

Bourguignon, F., M. Fournier, and M. Gurgand, "Selection bias corrections based on the multinomial logit model: Monte Carlo comparisons," *Journal of Economic Surveys* (21): 174–205, 2007.

Bradlow, E., "Encouragement Designs: An Approach to Self-Selected Samples in an Experimental Design," *Marketing Letters* (9): 383–391, 1998.

Braithwait, S., "Residential TOU Price Response in the Presence of Interactive Communication Equipment," in *Pricing in Competitive Electricity Markets.* A. Faruqui and K. Eakin (editors), Kluwer Academic Publishers, Boston, MA: 2000.

Brookhart, M. A., J. Rassen, and S. Schneeweiss, "Instrumental variable methods in comparative safety and effectiveness research," *Pharmacoepidemiology and Drug Safety* (19): 537–554, 2010.

Caves, D. and L. Christensen, "Global Properties of Flexible Functional Forms," *American Economic Review* (70): 422–432, 1980a.

Caves, D. and L. Christensen, "Residential Substitution of Off-Peak for Peak Electricity Usage Under Time-of-Use Prices," *Energy Journal* (1): 85–142, 1980b.

Caves, D. and L. Christensen, "Econometric Analysis of Residential Time-of-Use Pricing Experiments," *Journal of Econometrics* (14): 287–306, 1980c.

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

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

Caves, D., L. Christensen, P. Schoech, and W. Hendricks, "A Comparison of Different Methodologies in a Case Study of Residential Time-of-Use Electricity Pricing: Cost-Benefit Analysis," *Journal of Econometrics* (26): 17–34, 1984c.

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

Chambers, R., *Applied Production Analysis*. Cambridge University Press, Cambridge: 1988.

Charles River Associates, *Impact Evaluation of the California Statewide Pricing Pilot (Residential Summary)*. Charles River Associates, Oakland, CA, March 16, 2005.

Clogg, C., E. Petkova, and A. Haritou, "Statistical Methods for Comparing Regression Coefficients Between Models," *American Journal of Sociology* (100): 1261–1293, 1995.

Cornes, R., *Duality and Modern Economics*. Cambridge University Press, Cambridge: 1992.

Crepon, C., F. Devoto, E. Duflo, and W. Pariente, "Impact of Microcredit in Rural Areas of Morocco: Evidence from a Randomized Evaluation," Working Paper, available at: http://economics.mit.edu/faculty/eduflo/papers.

Davidson, R. and J. Mackinnon, *Estimation and Inference in Econometrics*. Oxford University Press, New York, NY: 1993.

Davis, L. and J. Boomhower, "Free Riders and the High Cost of Energy-Efficiency Subsidies," Working paper, University of California, Berkeley, 2013.

Deaton, A. and J. Muellbauer, *Economics and Customer Behavior*. Cambridge University Press, New York, NY: 1980.

Diewert, W., "An Application of the Shepard Duality Theorem: A Generalized Linear Production Function," *Journal of Political Economy* (79): 482–507, 1971.

Duban, J. and D. McFadden, "An Econometric Analysis of Residential Electric Appliance Holding and Consumption," *Econometrica* (52): 345–62, 1984.

Duflo, E., R. Glennerster, and M. Kremer, "Using Randomization in Development Economics Research: A Toolkit," *Handbook of Development Economics* (4): 3895–3962, 2007.

eMeter Consulting, *PowerCents DC™ Program Final Report*, prepared by eMeter Strategic Consulting for Smart Meter Pilot Program, Inc., 2010.

EPRI, 2008. *Characterizing and Quantifying the Societal Benefits Attributable to Smart Metering Investments*. EPRI, Palo Alto, CA: 2008. 1017006.

EPRI, 2010. *Methodological Approach for Estimating the Benefits and Costs of Smart Grid Demonstration Projects*. EPRI, Palo Alto, CA: 2010.1020342.

EPRI, 2010. *Guidelines for Designing Effective Energy Information Feedback Pilots: Research Protocols*. EPRI, Palo Alto, CA: 2010.1020855.

EPRI, 2010. *ComEd Customer Applications Program—Objectives, Research Design, and Implementation Details*. EPRI, Palo Alto, CA: 2010.1022266.

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

EPRI, 2011b. *The Effect on Electricity Consumption of the Commonwealth Edison Customer Application Program Pilot: Phase 1, Appendices.* EPRI, Palo Alto, CA: 2011. 1022761.

EPRI, 2011. *System for Understanding Retail Electric Rate Structures.* EPRI, Palo Alto, CA: 2011. 1021962.

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

EPRI, 2012a. *The Effect on Electricity Consumption of the Commonwealth Edison Customer Application Program: Phase 2 Supplemental Information.* EPRI, Palo Alto, CA: 2012. 1024865.

EPRI, 2012b. *Understanding Electric Utility Customers: What We Know and What We Need to Know*. EPRI, Palo Alto, CA: 2012. 1023562.

EPRI, 2012. *Appendices: Understanding Electric Utility Customers: What We Know and What We Need to Know*. EPRI, Palo Alto, CA: 2012. 1024867.

Freeman, Sullivan & Co., *2011 Load Impact Evaluation for Pacific Gas and Electric Company's SmartAC Program*," San Francisco, CA: April 1, 2012.

Freeman, Sullivan & Co., *2010 Load Impact Evaluation of San Diego Gas & Electric Company's Summer Saver Program,* San Francisco, CA: April 1, 2011.

Freeman, Sullivan & Co. and Charles River Associates, *California's Statewide Pricing Pilot (Commercial and Industrial Analysis Update)*. Freeman, Sullivan & Co. and Charles River Associates, Oakland, CA: June 28, 2006.

Faruqui, A. and S. Sergici, *Baltimore Gas & Electric Smart Energy Pricing Pilot - Summer 2008. BGE's Smart Energy Pricing Pilot, Summer 2008 Impact.* The Brattle Group, Inc., April 28, 2009.

George, S. and J. Bode, *2008 Ex Post Load Impact Evaluation for Pacific Gas and Electric Company's SmartRate™ Tariff*. Freeman, Sullivan & Co. San Francisco, CA: December 30, 2008.

Greene, W., *Econometric Analysis*, 5th ed. Prentice Hall, Upper Saddle River, NJ: 2003.

Greene, W., *Econometric Analysis*, 7th ed. Prentice Hall, Upper Saddle River, NJ: 2011.

Gujarati, D., "Use of Dummy Variables in Testing for Equality between Sets of Coefficients in Two Linear Regressions: A Note," *The American Statistician* (24): 50–52, 1970a.

Gujarati, D., "Use of Dummy Variables in Testing for Equality Between Sets of Coefficients in Linear Regressions: A Generalization," *The American Statistician* (24): 18–22, 1970b.

Hausman, J., "Specification Tests in Econometrics," *Econometrica* (46): 1251–1271, 1978.

Herriges, J., S. Baladi, D. Caves, and B. Neenan, "The Response of Industrial Customers to Electric Rates Based Upon Dynamic Marginal Costs," *Review of Economics and Statistics* (75): 446–454, 1993.

Imai, K., G. King, and E. Stuart, "Misunderstandings between Experimentalists and Observationalists about Causal Inference," *Journal of the Royal Statistical Society Series A* (171): 481–502, 2008.

Imbens, G., and J. Angrist, "Identification and estimation of local average treatment effects," *Econometrica* (62): 467–475, 1994.

Imbens, G. W. and T. Lemieux, "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics* 142 (2): 615–635, 2008.

Imbens, G. and J. Wooldridge, “Recent Developments in the Econometrics of Program Evaluation,” *Journal of Economic Literature* (47): 5–86, 2009.

Ito, K., “Does Conservation Targeting Work? Evidence from a Statewide Electricity Rebate Program in California,” Unpublished manuscript. Stanford University, 2012.

Jadad, A., *Randomised Controlled Trials: A User’s Guide*. BMJ Books, London: 1998.

Jaskow, P., “Creating a Smarter U.S. Electricity Grid,” *Journal of Economic Perspectives* (26): 29–48, 2012.

Kennedy, P., *A Guide to Econometrics 6th Edition*. Blackwell Publishing, Malden, MA: 2008.

LaFrance, J., The Structure of Constant Elasticity Demand Models,” *American Journal of Agricultural Economics* (68): 541–52, 1986.

Marschak, J., “Binary Choice Constraints on Random Utility Indicators,” In K. Arrow, ed., *Stanford symposium on mathematical methods in the social sciences*. Stanford University Press, Stanford, CA: pp. 321–29, 1960.

McFadden, D., “Economic Choices,” *American Economic Review* (91): 351–78, 2001.

McFadden, D., “Regression-Based Specification Tests for the Multinomial Logit Model,” *Journal of Econometrics* (34): 63–82, 1987.

Neenan, B., D. Pratt, P. Cappers, R. Boisvert, and K. Deal, “NYISO Price-Responsive Load Program Evaluation Final Report,” Prepared for New York Independent System Operator, Albany, NY, 2002a.

Neenan, B., R. Boisvert, and P. Cappers, “What Makes a Customer Price Responsive?” *The Electricity Journal* 15(3): 52–59, 2002b.

Neenan, B., D. Pratt, P. Cappers, J. Doane, J. Anderson, R. Boisvert, C. Goldman, O. Sezgen, G. Barbose, R. Bharvirkar, M. Kintner-Meyer, S. Shankle, and D. Bates, *How and Why Customers Respond to Electricity Price Variability: A Study of NYISO and NYSERDA 2002 PRL Program Performance.* Report to the New York Independent System Operator (NYISO) and New York State Energy Research and Development Agency (NYSERDA), January 2003.

Paternoster, R., R. Brame, P. Mazerolle, and A. Piquero, “Using the Correct Statistical Test For The Equality Of Regression Coefficients,” *Criminology* (36): 859–866, 1998.

Patrick, R., “Rate Structure Effects and Regression Parameter Instability Across Time-of-Use Electricity Pricing Experiments,” *Resources and Energy* 12(2): 179–195, 1990.

Patrick, R. and F. Wolak, *Estimating the Customer-Level Demand for Electricity Under Real-Time Market Prices.* National Bureau of Economic Research (NBER) Working Paper 8213. April 2001.

Pollak, R., "Conditional Demand Functions and the Implications of Separable Utility," *The Southern Economic Journal* 37(4): 423–433, 1971.

Reiss, P. and M. White, "Household Electricity Demand, Revisited," *The Review of Economic Studies* (72): 853–883, 2005.

Results of CL&P's Plan-It Wise Energy Pilot. Connecticut Light and Power, Filing in Response to the Department of Public Utility Control's Compliance Order No. 4, Docket No. 05-10-03RE01, December 2009.

Silberberg, E. and W. Suen, *The Structure of Economics: A Mathematical Analysis*. 3rd Ed. Irwin/McGraw-Hill, Boston, MA: 2001.

Smith, D. and R. Paternoster, "The Gender Gap in Theories of Deviance: Issues and Evidence," *Journal of Research in Crime and Delinquency* (24): 140–172, 1987.

Strong, A. and V. Kerry Smith, "Reconsidering the Economics of Demand Analysis with Kinked Budget Constraints," *Land Economics* (86): 173–90, 2010.

Taylor, T. and P. Schwarz, "The Long-Run Effects of a Time-of-Use Demand Charge," *The Rand Journal of Economics* (21): 431–445, 1990.

Taylor, T., P. Schwarz, and J. Cochell, "24/7 Hourly Response to Electricity Prices: Pricing with up to Eight Summer's Experience," *Journal of Regulatory Economics* (27): 235–262, 2005.

Thurstone, L., "A Law of Comparative Judgement," *Psychological Review* (34): 273–86, 1927.

U.S. Department of Energy. U.S. Department of Energy Smart Grid Investment Grant Technical Advisory Group Guidance Document #7: Design and Implementation of Program Evaluations that Utilize Randomized Experimental Approaches. November 2010.

U.S. Department of Energy. Guidebook for ARRA Smart Grid Program Metrics and Benefits. 2009.

U.S. Department of Energy, Smart Grid Investment Grant Technical Advisory Group. Guidance Document #2, Topic: Non-Rate Treatments in Consumer Behavior Study Design. 2010a.

U.S. Department of Energy, Smart Grid Investment Grant Technical Advisory Group. Guidance Document #3, Topic: Use of Stratification and Sample Weights for Smart Grid Demonstration Projects Using Experimental Data. 2010b.

U.S. Department of Energy, Smart Grid Investment Grant Technical Advisory Group. Guidance Document #4, Topic: Rate Design Treatments in Consumer Behavior Study Design. 2010c.

U.S. Department of Energy, Smart Grid Investment Grant Technical Advisory Group. Guidance Document #5, Topic: Techniques for Estimating Impact Measurements. 2010d.

U.S. Department of Energy, Smart Grid Investment Grant Technical Advisory Group. Guidance Document #6, Topic: Recommendations for Content of the Consumer Behavior Study Evaluation Report(s). 2010e.

U.S. Department of Energy, Smart Grid Investment Grant Technical Advisory Group. Guidance Document #9, Topic: Preferences for DOE Required Data Collection via Survey Instrument. 2011.

Wooldridge, J. *Introductory Econometric: A Modern Approach* (4th ed.). South-Western Cengage Learning, Mason, OH: 2009.

Wooldridge, J. *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). The MIT Press, Cambridge, MA: 2010.

Zelen, M. "A new design for randomized clinical trials," *New England Journal of Medicine* (300): 1242–1245, 1979.

7. Appendix A:Panel Data Fixed-Effects Estimators

A data set consisting of hourly load data by customer, for many customers, has both cross-sectional and time-series components that define a panel data structure. In other words, panel data provide measurements of certain characteristics (load, in this case) for many of the same subjects (customers) over many time periods (hours over the period of the pilot study). Comparisons across customers in any time period involve a cross-sectional analysis. Comparisons of single customers over time involve a time-series analysis. Panel data methods, in contrast, include both perspectives and provide a way to isolate treatment effects among customers.

It follows that load impacts for CBS pilots can and should be modeled using methods designed specifically for the analysis of data that have this particular panel data structure.[53] As suggested in the DOE's TAG guide, *Techniques for Estimating Impact Measurements*, panel data models can be structured empirically to estimate the effects of the three primary types of factors that affect customers' electricity consumption:

1. Factors that are fixed for the customer over the study time period but that differ across customers (for example, household income, dwelling type and size, and appliance stock)
2. Factors that do not vary across customers at any one time but that do differ over time (for example, day of the week or hour of the day)
3. Factors that differ over time but through interactions with factors that are fixed for the customer (for example, weather and central air conditioning usage)

The fixed factors (Item 1) are accounted for in the way in which the model is structured and estimation is accomplished, but individual customer effects are seldom explicitly derived. By accounting for the effects of the factors that differ over time, panel regression models can potentially explain much more variation in electricity consumption than other techniques that do not account for the time variation in such factors. Because more of the variation in patterns of electricity usage is potentially explained in the panel regression model, it is also likely that the estimates of treatment impacts will be measured with greater precision. Moreover, these models may correct for bias in the treatment effects by controlling for the effects of otherwise omitted variables—the constant customer effects.

Put somewhat differently, most researchers would probably agree that cross-sectional heterogeneity reflected in these panel data is likely to be the norm. That is, there are unmeasured variables that, at least in part, determine each customer's electricity consumption; one should account for their combined influence. This influence can be captured by specifying a different customer effect in a regression model used to explain differences in electricity consumption across customers.

---

[53] Wooldridge (2010) argues that for panel data, which consist of repeated observations on the same cross sections of individuals, households, or firms over time, the random sampling assumption appears to be too restrictive. There ought to be allowance for correlation in firm or household behavior over time. In the analysis of panel data, particularly when there are many observations in the cross section over a relatively short period of time, it is often assumed that there is random sampling in the cross section but with unrestricted dependence over time.

In the econometrics literature, two major types of panel data models have been suggested to improve the estimation when faced with this unobservable heterogeneity: a fixed-effects model and a random effects model. Each involves a different approach to accounting for the presence of a different intercept for each customer (that is, a cross-sectional unit).

According to Allison (2005), Wooldridge explains that in a modern econometric framework, the unobserved differences (in this case, among customers) are always regarded as random variables. The two models differ in what is assumed about the structure of the correlations between the variables that are observed and those that are not. In a random effects model, the unobserved variables are assumed to be uncorrelated with observed variables. In fixed-effects models, the unobserved variables can have any correlation with the observed variables. The latter specification makes it convenient to account for effects of unobserved variables; it is equivalent to treating the unobserved variables as fixed parameters embodied in the constant term of the specification.

For this reason, the fixed-effects model is also equivalent to a model in which these influences are captured by the presence of a separate intercept for each customer through the inclusion of a separate dummy variable (which defines the constant of the recession) for each customer. In this way, the individual customer effects are measured by the coefficients on customer-specific dummy variables in an OLS regression. Although this makes sense conceptually, such an application would be problematic because it would require the introduction of a dummy variable into the model for each of the customers. For short time periods and with many customers (which reduces degrees of freedom in estimation), this may make estimation impossible or erode the measures of statistical significance.

This difficulty is avoided through the application of the fixed-effects estimator, which, as Kennedy (2010) remarks, is OLS applied to the fixed-effects model. The fixed-effects estimator avoids the difficulty in the degrees of freedom through a transformation of the data, accomplished by subtracting from each observation (measured variables) for each customer the average of the observations for that customer. In so doing, OLS can be applied to the transformed data to obtain the desired parameter estimates.[54]

Kennedy (2008) also points to another major drawback to the fixed-effects model, and it is potentially more serious. The transformation of the data involved in the estimation process effectively fails to take into account the effects of explanatory variables that do not vary over time for an individual customer. This would seem to preclude adding as independent variables in the model characteristics that are constant by customer and may have an effect, such as income, household size, and premise central air conditioning. This happens because the values of these variables have the same value for the customer in every time period. Therefore, when the values for each customer for each time period are subtracted from the average across time periods, the variables take on a value of zero.

[54] This fixed-effects estimator or its equivalent is contained in many widely used econometrics software packages. In estimating, many apply procedures such as generalized least squares (GLS) to control for heteroskedasticity.

Kennedy (2010) goes on to explain that the two drawbacks can be overcome through the application of a random effects model in which there is a separate intercept term for each customer, but the intercepts are viewed as having been drawn from a pool of possible values. The intercept terms are interpreted as random—included in and accounted for in the error term rather than the constant of the regression.

The random effects model seems to account for independent variables in the model that do not vary over time by customer. However, to test for whether the fixed- or random effects model is the most appropriate specification, it is necessary to test whether the random effects estimator is unbiased.[55] Moreover, because random effects models embody the assumption that all of the variables omitted from the regression function have random influence on the dependent variable, they are assumed to be uncorrelated with any of the independent variables included in the regression. This is a strong assumption; it is justified only when observations have been randomly assigned to treatment and control conditions, with no significant attrition in treatment or control groups, which defines an RCT. Because these conditions are difficult to achieve in practice, this random effects model is rarely used (DOE 2010d).

If the random effects estimator is not unbiased (and, in practice, this would seem likely to be the case), the fixed-effects model is appropriate. Moreover, we are again faced with the conundrum that we cannot identify the effects of an explanatory variable that is time-invariant for each customer when there is no variation to explain. It turns out that the solution to this dilemma is rather simple, but only if one is willing to assume that customer or premise characteristics (for example, air conditioning) affect usage primarily through their interaction with treatments and/or factors (for example, weather) that do vary over time. These effects are captured by coefficient associated with variables created with interaction terms between these types of variables and those that represent the various treatments.

The Fixed-Effects Model

This fixed-effects model is being employed in the analysis of dynamic rate pilots, particularly to estimate such performance measures as the peak load impacts during CPP events (EPRI 1024867). Although the specific empirical specifications differ across studies, the fact remains that the approach is equivalent to including a separate dummy (0, 1) variable for each customer.[56]

[55] Kennedy (2010) suggests the following strategy for estimation with panel data. The process begins by testing that the null hypothesis that the customer-specific intercepts are equal. If one fails to reject this null hypothesis, the data can be pooled. If this null hypothesis, however, is rejected, a Hausman test is applied to test whether the random effects estimator is unbiased. If one fails to reject this test, the random effects estimator is used. If the null hypothesis is rejected, the fixed-effects estimator is used.

[56] Wooldridge (2010) also contains an excellent discussion of fixed-effects models. He emphasizes that the econometrics software that computes the fixed effects rarely reports the "estimates of the individual customer effects" (at least in part because there are typically a very large number of them). He goes on to argue that because these effects are equivalent to the coefficients on customer-specific dummy variables in an OLS regression, it is possible to compute the sample average, sample median, or sample quartiles of these effects to gain some idea of the way in which the heterogeneity is distributed across the sample.

In its simplest form, this model would include only those independent dummy variables associated with the treatment groups. Because the different customer intercepts control for effects unique to that customer but constant over the time period, some analysts would seem to argue that because of its fixed-effects nature, the model does not need to include unchanging customer characteristics such as premise square footage, stock of electrical appliances, and household socioeconomic characteristics. Because each customer has a different base load, a different response to weather, and different patterns of consumption that change over time, the inclusion of these fixed-effects controls for the amount of variance in the model. By including fixed (stationary) time effects, the model would also control for differences in consumption across days as a result of temperature, sunshine, and any other variable over time factors common to all customers for any day.

It is likely that such a model could explain a large portion of the variation in customer load. However, because the only included independent variables are for the rate treatment effects and the event days, this type of application of fixed-effects approach for measuring load impacts is extremely compact and stylized. Although computationally convenient, this formulation is shortsighted in that it fails to investigate which exogenous factors (those outside of the control of the rate provider) are exerting an influence on customer usage and to measure the size of that influence. That influence (or lack thereof) may be the result of particular market conditions in which the study was conducted to specific customer and premise characteristics. Failure to account for the effects of these factors certainly limits the usefulness of the results, especially in other contexts. Without knowing how customer and premise characteristics affect usage, it is impossible to use the results to market alternative rates or, for impact distributional considerations, to know the effects of rates on socioeconomic groups, such as the elderly or those with low incomes.

This shortcoming of the stylized specification of the fixed-effects model, however, does not render the fixed-effects structure inappropriate. Rather, we argue that the model should be constructed to include the variables necessary to account for the rate treatment effects, the effects of prices or price ratios, and differences in important factors such as customer demographics, premise characteristics, and weather.

As explained previously, the added complexity can be accommodated within this fixed-effects structure if one is willing to assume that these customer characteristics (for example, air conditioning) affect usage primarily through their interaction with treatments and/or factors (for example, weather) that vary over time. Because this would seem to be a reasonable hypothesis to test, the effects of these factors can be captured by the coefficients associated with variables created by constructing interaction terms between these types of variables with those that represent the various treatments. Furthermore, by constructing the model in this way, we can easily add interaction terms to identify the way in which the rate treatments interact with enabling technology and socioeconomic and other identifiable customer characteristics to determine load impacts. Within an expanded fixed-effects model, the customer fixed effects still account for remaining unobserved factors that result in customer heterogeneity.[57]

[57] In Section 4, we discuss models to estimate two measures of price response: an own-price elasticity and an elasticity of substitution. These models have been estimated using a panel data fixed-effect estimator. For example, the analysts who conducted the evaluation of the recent Baltimore Gas & Electric Smart Energy Pricing Pilot

A balanced modeling strategy has the potential to identify or decompose the separate effects of these important variables that affect customer load impacts—and the results are important for marketing and for understanding important policy implications of alternative rate structures. However, the benefits from this more balanced modeling strategy also underscore the need to obtain reliable customer-level information that can be combined with the customer load and price data. We cannot stress enough the importance of having access to these additional data for most—and better still, all—customers in the sample.[58] The survey questionnaire must be carefully designed to collect what might be the important influences, and there must be a concerted effort to administer the survey in a way that ensures a high customer response rate.

(Faruqui and Sergici 2009) adopt what we have called here a more *balanced* approach. They specify a two-equation demand system to estimate an own-price elasticity and an elasticity of substitution. In each of the demand equations, they include the customer fixed effects to control for customer-specific characteristics that do not vary over time. The differences in the price effects across treatments can still be estimated by including interaction terms between the appropriate price variables and the dummy variables for the treatments, even though the treatments do not vary over time for any customer.
The analysts of the recent assessment of PG&E's Smart Rate Tariff™ employ an alternative to the fixed-effect approach (George and Bode 2008). They estimate time series regressions at the individual customer level rather than pooling the data for all customers. In so doing, these analysts argue that they not only capture the effects particular to the unobservable characteristics of each customer, but they also allow for the coefficients on the other variables in the model to differ by customer. They argue that this strategy leads to results that are more accurate at the customer level and that it facilitates the calculation of the effects by customer segments in addition to the average for all participants. The strategy does come at the cost of substantial additional computational time. If there are data from survey responses on the socioeconomic characteristics of customers and characteristics of premises, one might also exploit the richness of the panel data fixed-effects model and still generate many of the same effects using interaction terms between treatments, prices, and customer characteristics.
58 If survey data are collected for only some of the subjects, the data available for this more comprehensive assessment of effects are limited to customers for whom there are data. The results may then be valid only for the subset of customers, and the ability to generalize to a broader population may be extremely limited.

8. Appendix B: Analysis of Variance and/or Covariance

Because the experimental designs of most pilot studies involve a series of treatment and control groups, one approach to identify treatment effects is to apply methods of analysis of variance (ANOVA) to test for differences in the average electricity consumption of various types (for example, average daily consumption and average hourly peak-period consumption) between treatment and control groups.

ANOVA is a statistical technique designed to determine whether a particular classification of the data is meaningful, such as whether the mean performance measures of impacts differ among the various treatment and control groups. The ANOVA test is based on the decomposition of the total variation in the dependent variable (the measure of performance) into that resulting from the variation between treatments and that from the variation within each treatment. This decomposition is used to construct an *F* statistic to test the hypothesis that the between-treatment variation is large relative to the within-treatment variation. This result assumes that the treatment classification is meaningful and that there is significant variation in the dependent variable between treatments.

An ANOVA test can be conducted in a convenient way through the specification of a dummy variable regression model, in which the dependent variable is the impact measure by customer and the independent variables are (0, 1) dummy variables for each treatment. These variables are assigned a value of unity if the customer belongs to that treatment group, and zero otherwise. As Kennedy (2008) points out, the coefficients on the dummy variables are the treatment means, the between-treatment variation is the regression's explained variation, the within-treatment variation is the regression's unexplained variation, and the ANOVA *F* test is equivalent to testing whether the dummy variable coefficients are significantly different from one another. Kennedy (2008) further states: "The main advantage of the dummy variable regression approach is that it provides estimates of the magnitudes of the [treatment] variations on the dependent variables (as well as testing whether the classification is meaningful)."

Within a regression framework, the standard ANOVA test is a joint test that all of the regression coefficients for all treatments are not different from zero. Anderson et al. (1999), for example, provide a simple demonstration that one can also generate these standard ANOVA results from a linear regression model in which the explanatory variables are dummy (0, 1) for n-1 treatments. The test of the joint null hypothesis that the means of the performance measure (for example, the dependent variable) are no different across treatments is an *F* test for the regression line.
For example, assume that the model has *M* customers, each assigned to one of *N* treatments. If, for simplicity, there are three treatments (that is, N = 3), we can say that Treatment 3 is the control group. We define two (0, 1) variables for two of the treatments; recall that we need only define N-1 dummy variables. The regression model is as follows:

Eq. B-1: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$

where Y_i *is* electricity use during some time period for customer *i*, X_{1i} is 1 if customer *i* is in Treatment Group 1 and zero otherwise; X_{2i} is 1 if customer *i* is in Treatment Group 2 and zero otherwise; and ε_i is an error term. The coefficient β_0 is the expected value of *Y* for those customers in Treatment 3, perhaps the control group, while $\beta_0 + \beta_1$ is the expected (mean) value of *Y* for customers in Treatment Group 1, and $\beta_0 + \beta_2$ is the expected (mean) value of *Y* for Treatment Group 2.

If, as in ANOVA, we want to test the null hypothesis that there are no differences in the means of *Y* across the groups, this is equivalent to a joint test of H_0: $\beta_1 = \beta_2 = 0$. This is a test of the significance of the regression relationship itself. To perform the test, we compute the ratio of the mean between-group sum of squared errors and the mean within-group sum of squared errors. This ratio follows an *F* distribution with degrees of freedom (N-1, M-2). If one fails to reject the null hypothesis, it is unlikely that we would reject the null hypotheses for both of the individual coefficients on a treatment variable based on the standard t-tests of individual regression coefficients.

Kennedy (2008) goes on to explain that the analysis of covariance (ANCOVA) is an extension of ANOVA designed to handle situations in which there are some uncontrolled variables that cannot be standardized between treatments. Although these situations can still be modeled using dummy variables to capture the treatment classification, in this case we regress the dependent (performance) variable on both the treatment dummy variables and the other uncontrolled variables. In so doing, the *F* test for the ANCOVA is equivalent to testing whether the coefficients of all of the dummy variables are statistically significantly different from one another.

By constructing the models in this way, the observable factors that cannot be standardized between treatments can be used to explain some of the variation in electricity consumption within any particular cross section. These effects are in addition to the effects of the (0, 1) variables that account for the presence or absence of a particular treatment. Through the construction of interaction terms between these additional variables and the treatment indicator variables, the regression equations can include unique slope parameter(s) as well as intercept parameters for the treatment and control groups. If any of the slope parameters for the treatment and control groups is not significantly different from zero, the differences in the intercepts can be interpreted as the treatment effects. If this is not the case, one must conclude that the impacts of the treatments are affected by the interaction between the treatments and the other observable factors (for example, household or premise characteristics) because the slopes—as well as the intercepts for the two groups—are different.

The incorporation of covariates related to customer circumstances to control for some cross-sectional variation in the performance measures may improve the efficiency of the estimation of the impact on electricity consumption, in particular if the other factors are influential. However, it can also be argued that that even these ANCOVA models still fail to take advantage of the efficiency gains in estimation made possible by using the additional data and information being collected as a result of the panel structure of the data (for example, DOE 2010d). By using this additional information, one may capture the effects of variables that change over time. The ability to exploit the panel structure has been enhanced dramatically in recent years as a result of the advances in our understanding of the econometrics of panel data and the development of software that facilitates the estimation of these panel data models.

9. Appendix C: Measuring Treatment Effects – Difference-in-Differences Estimator

Under ideal conditions in conducting rate pilots, we will have access to hourly load data for all customers in the rate treatments as well as similar data for all customers in a valid control group. Furthermore, if the hourly load data for both treatment and control groups are for time periods both before and after the beginning of the treatment, we can use the pooled cross-sectional data to isolate and quantify the effects of these rate treatments on one or more measures of load response as mentioned above.

In the econometrics literature, models that exploit the richness of the data contained in pooled cross-sectional data sets with these characteristics are often discussed within the general context of policy analysis, and the focus in on what has become known as a *difference-in-differences estimator*.[59] In unpublished lecture notes, Imbens and Wooldridge state that the simplest set up is one "where outcomes are observed for two groups for two time periods. One of the groups is exposed to a treatment in the second period but not in the first period. The second group is not exposed to the treatment during either period. In the case where the same units within a group are observed in each time period, the average gain in the second (control) group is subtracted from the average gain in the first (treatment) group."

In our case, we would be interested in the differences in the energy usage or demand for electricity "(before and after treatment) between consumers who have been randomly assigned to a given combination of pricing and/or information and automation/control technology and a control group that was not exposed to the combination" (DOE 2010d). If we let Y_{gti} equal our measure of electricity consumption for customer $i = 1,\ldots, I$ in group $g = c$ (control), m (treatment) and time $t = 1$ (pre-treatment), 2 (post-treatment), the difference-in-differences (D-in-D) measure of the impact of the treatment can be given by Eq. C-1:

Eq. C-1: $\widehat{impact}_{d-in-d} = \left(\bar{Y}_{m,2} - \bar{Y}_{m,1}\right) - \left(\bar{Y}_{c,2} - \bar{Y}_{c,1}\right)$

where $\bar{Y}_{g,t}$ is the average electricity consumption for individuals in group g and time period t. We simply compute the changes in the average consumption over time for each of the two groups and then we calculate the difference between these changes. Because the order in which the differences are taken does not matter, we could alternatively compute the differences in the averages between the two groups in each time period and then calculate the difference over time. This leads to an equivalent expression for the estimated impact:

[59] The methodology applied to before and after the treatment panel data sets also has numerous applications when the data are generated from some natural or quasi-experiment. Wooldridge (2006) suggests that a "natural experiment occurs when some exogenous event—often a change in government policy—changes the environment in which individuals, families, firms, or cities operate" (p. 458). Therefore, a natural experiment always has a control group (for example, those not affected by the policy change or other exogenous event) and a treatment group (for example, those affected by the change in policy or the exogenous event). However, in the case of a natural experiment, the control and treatment groups are determined by the nature of the exogenous event or policy rather than being randomly and explicitly chosen, as in a true experiment.

Eq. C-2: $\widehat{impact}_{d-in-d} = (\bar{Y}_{m,2} - \bar{Y}_{c,2}) - (\bar{Y}_{t,1} - \bar{Y}_{c,1})$

The null hypothesis that there are no differences in electricity consumption between the two groups can be tested using a simple t-test or through ANOVA. Although this D-in-D estimator is easy to calculate, it is unlikely to be of much use in this simple form for studying changes in electricity consumption. It suffers from the same limitations as other ANOVA analyses because electricity consumption differs greatly among randomly selected consumers. These differences are the result of customer and premise characteristics such as dwelling size, household size, occupancy patterns, and appliance holdings.

This D-in-D estimator can also be developed within a linear regression framework; we can demonstrate how this model can be expanded to overcome several of its limitations. The regression model is for any generic customer in any of the groups:

Eq. C-3: $Y = \alpha_0 + \delta_0 D_2 + \alpha_1 D_m + \delta_1 D_2 D_m + \varepsilon$

where D_2 is a dummy variable that takes on a value of unity if the observation is from the second (post-treatment) period and zero otherwise, and D_m is a dummy variable that takes on a value of unity if the observation is for a member of the treatment group and zero otherwise. Therefore, the treatment variable D_m captures any differences in consumption between the treatment and control groups, while δ_0, the coefficient on the time period variable D_2, captures the effect on Y that would have occurred even without the rate treatment. Finally, the coefficient of interest is δ_1 because it defines the interaction between the time period and the treatment dummies, $D_2 D_m$, which is the same as a dummy variable equal to unity for those post-treatment observations for customers in the treatment group. Therefore, δ_1 is often called the *average treatment effect* because it measures the effect of the treatment on the average outcome of Y. Because the coefficients on dummy variables in regression models capture differences in mean outcomes for a particular group, this coefficient is equivalent to the measure in Eq. C-1 and C-2. The regression model in Eq. C-3 is often decomposed as in Table C-1, which measures before and after the treatment for both the control and treatment groups and for the differences.

Table C-1: Before and After the Treatment for Control and Treatment Groups

	Before	After	After - Before
Control	α_0	$\alpha_0+\delta_0$	δ_0
Treatment	$\alpha_0+\alpha_1$	$\alpha_0+\delta_0+\alpha_1+\delta_1$	$\delta_0+\delta_1$
Treatment – Control	α_1	$\alpha_1+\delta_1$	δ_1

When viewed as a linear regression model, it is easy to envision how this model that generates the D-in-D estimator can be extended to account for differences in electricity consumption among randomly selected consumers resulting from customer and premise characteristics such as dwelling size, household size, occupancy patterns, and appliance holdings. The strategy is simply one of adding covariates to the model that reflect these differences in customer and premise characteristics. This more general model is now as shown in Eq. C-4:

Eq. C-4: $Y = \alpha_0 + \delta_0 D_2 + \alpha_1 D_m + \delta_1 D_2 D_m + \sum_{j=1}^{J} \gamma_j F_j + \varepsilon$

where F_j is the j^{th} customer or premise characteristic (for example, dwelling size, household size, occupancy patterns, or appliance holdings) for a representative member of any group, and other variables are defined as above. The only caution in interpreting the results of this more robust model is that the OLS estimates of δ_I no longer has the simple form as depicted in Eq. C-1 or C-2, but its interpretation is similar (Wooldridge 2006, p. 459).

10. Appendix D: Electricity Demand Models to Estimate Load Shifting

10.1 Conceptual Models for Electricity Demand

Our methodologies for estimating the effects of alternative electricity rate structures and enabling technologies on electricity usage are based on the neoclassical theory of customer behavior. As suggested by Caves and Christensen (1980b, c) and others in their analyses of early electricity pricing experiments, such an approach ensures that the empirical specification of the estimated demand equations is consistent with the maximization of customer utility (for example, satisfaction) subject to a budget constraint and that the estimated demand elasticities are internally consistent.

Further, by placing certain restrictions on the form of the utility function, it is possible to conceptualize the analysis in stages. This is important for several reasons. Residential customers purchase electricity along with a large number of commodities, including other types of energy and products such as housing, transportation, clothing, food, health care, education, and recreation. Within each of the major categories, expenditures are allocated among the subcomponents of each—for example, among meat, vegetables, and grains in the food category.

However, as a commodity, electricity is unlike most others. Because it is not storable, it is generally purchased continuously throughout the day on an as-needed basis. Equally important, electricity is not consumed directly by customers. Rather, its demand is a derived demand because customers derive satisfaction (for example, utility) from the services that come from electrical appliances. Therefore, the satisfaction from purchases of electricity is embodied in the derived demand for distinct services such as lighting, HVAC, electronic devices, ovens, and refrigerators. In the short term, a utility customer's demand for electricity is conditioned by an existing stock of electricity appliances. As suggested by much of the research into the value of feedback information, the demand for electricity may well be further conditioned by the availability of advanced metering infrastructure (AMI) technology and education (for example, Boisvert et al. 2009).

By viewing the analysis of electricity demand in stages, we can distinguish at a minimum between a customer's allocation of electricity purchases among different time periods from the allocation of total income or expenditures between electricity and other nondurable goods and services in any time period. Beyond the transparency it brings to the analysis, Caves and Christensen (1980b) argue that it is desirable, if not essential, to analyze the allocation of electricity expenditures in stages because of the availability of data.

As in other experiments with time-differentiated electricity rates, the pilot involving the five rate treatments in ComEd's CAP provides extensive data on electricity use and rates as well as some data on structure characteristics and appliance stock—but very little, if any, information on the purchase of others goods. At best, there are likely to be only crude estimates of income. By focusing exclusively on electricity use while disregarding income and the customer's total budget, we can make good use of the detailed data from the pilot. However, in so doing, our estimates of the important demand elasticities are only partial elasticities.

As explained in more detail next, implicit in this two-stage analysis is the assumption that electricity is separable in the customer's utility function.[60] Accordingly, the customer's budgeting process can be viewed as proceeding in two stages. Because we are especially interested in the incentives that dynamic rates provide for shifting load between high- priced and low-priced times of the day, much of our focus is on the first stage of the process in which we can characterize the allocation of electricity use by time-of-day. Although the elasticities from this stage are partial elasticities of substitution and partial price elasticities, they have several useful interpretations that are also discussed next.[61]

10.2 Conditional Demand for Electricity

In modeling the customer's allocation of daily electricity for several rate treatments that differ by hour, it is conceptually possible to model each hour of the day. For purposes of discussing the modeling framework, however, it is sufficient to examine the allocation of electricity consumption between high-priced (peak) hours and low-priced (off-peak) hours.[62] This issue is addressed in detail next.

Because this modeling framework measures the amount of load shifted from peak to off-peak periods, it is particularly appropriate to analyze rates that differ by time of day, regardless of whether the rate involves fixed time-of-use (TOU) rates or rates that have a dynamic aspect such as day-ahead real-time pricing (DA-RTP) or critical peak pricing (CPP). The model may also offer limited information about overall energy conservation.

---

[60] The critical assumption underlying the "separability" of consumption is that an individual's preference between two collections of goods that differ only in the components of one subset of a category are independent of the identical other components of another category. In addition to its intuitive appeal, this assumption of separability allows for the identification of conditional demand functions for goods in any category. These conditional demand functions can be defined for when one or more goods is pre-allocated. In general, a conditional demand function for a good in the remaining subset that is not pre-allocated expresses the demand for that good as a function of 1) the prices of all goods in the subset of goods not pre-allocated, 2) total expenditures on the subset of goods, and 3) the quantities of pre-allocated goods (for example, Pollak 1971, p. 424).

[61] The primary focus of empirical analysis of the demand for electricity is on estimating the elasticity of substitution between peak and off-peak electricity demand and how this differs for customers across rate treatments. This elasticity of substitution is often denoted by σ, and in our case it measures the substitution effect quantified by the percentage change in the ratio of peak to off-peak electricity use caused by a 1% change in the ratio of off-peak to peak electricity prices. In conducting the empirical analysis, we do, however, also obtain estimates of the conditional own-price elasticities of demand for peak and off-peak electricity. These are defined as the percentage changes in peak (off-peak) electricity use caused by a 1% change in the price of peak (off-peak) electricity. These own-price elasticities are estimated primarily to check that the estimated models are consistent with demand theory. They do offer some measure of demand response in a particular time period to changes in the price in that period, but they must be interpreted with care. Because we have no data on customers' income, our estimated own-price elasticities of demand for peak and off-peak elasticity are measured conditional on the level of a customer's utility remaining unchanged.

[62] The use of the terms *peak* and *off-peak* is primarily for convenience and distinguishes between hours in which prices are generally high and those when they are low. This division of hours may or may not correspond with the system peak.

In contrast, this model of the first-stage decision process is not sufficient to examine rates such as an inclining block rate (IBR) because prices under this rate do not vary by time of day. The price varies by the quantity consumed. The primary decision of customers under the IBR rate involves how much electricity to consume during the billing period rather than how much electricity to consume at different hours of the day. Through its inclining block, the IBR rate embodies an incentive to reduce overall electricity consumption; the critical part of the customer's choice is the selection of the block within which to consume the last unit of electricity. A discussion of the framework necessary to study customer price response under an IBR rate is beyond the scope of this appendix.

10.3 Modeling Customer Response to Prices that Differ by Time of Day

Electricity uses in two distinct daily periods (for example, peak and off-peak) may be valued differently by the customer. Peak and off-peak electricity consumption may also be complementary goods so that the customer demands electricity during the two periods in nearly fixed proportions (for example, Taylor et al. 2005 and Boisvert et al. 2007). To capture these ideas in a demand model, we specify a customer's utility function that is separable in electricity commodities as:

$$V = V\left(x_1, x_2, ..., x_n, U\left(k_p, k_o\right)\right)$$

where V is the utility function of the customer, x_i is the goods and services other than electricity consumed, and k_p and k_o are the amounts of electricity consumed in peak and off-peak periods, respectively. Electricity is assumed to be separable in consumption from other goods and services. Therefore, the sub-function $U(k_p,k_o)$ represents a sub-utility function for the customer; it reflects the fact that a customer can attain a given level of satisfaction from electricity consumption by consuming different amounts of peak and off-peak electricity that together yield a given level of utility or satisfaction, say U_0.

In studying the impact of dynamic rates, it is important to measure the customer's load shifting from peak (or event) hours to off-peak (non-event) hours. This is accomplished by estimating an *elasticity of substitution*, defined as the percentage change in the ratio of peak to off-peak usage resulting from a 1% change in the ratio of off-peak to peak prices. Because this change in the ratio of usage is related to the change in the inverse price ratio, the sign of the elasticity of substitution is expected to be positive.

The estimated elasticity of substitution is consistent with conventional demand models for electricity, and there are two empirical specifications that dominate the empirical studies. One is based on a constant elasticity of substitution (CES) demand model, and (as its name suggests) the estimated elasticities of substitution are constant—they do not depend on the nominal levels of prices.[63]

In contrast, the elasticity of substitution can also be estimated based on the application of the generalized Leontief (GL) demand model. This model is much more flexible than the CES model, and through its application the estimated elasticities of substitution can differ depending on the nominal prices of peak and off-peak electricity. Although this flexibility may be more important in the study of some dynamic rate designs than others and more critical for some

[63] As indicated below, the elasticities of substitution based on the CES model can, however, differ based on weather or customer or premise characteristics.

customer classes than others, its application is not without its risks: this flexible-form demand function imposes virtually no structure on the underlying demand system. Based on the GL specification, we can control for each customer's typical load profile and for the effects of weather. These customer-specific regressions provide estimates of the elasticities of substitution that differ by day.

Although this GL model has some attractive features and accommodates differences in elasticities of substitution by customer and by day, the computational burden in its application, relative to the CES model, is substantial. Therefore, before adopting this approach, it is important to determine whether such a flexible form is needed and that the computational complexities are therefore justified.

It is primarily for this reason that we begin this appendix with a discussion of the CES model. The major limitation of this model in its simplest form is that the elasticities of substitution are constant at all data points. However, in any demand model specification, it is important to include some non-economic variables, such as those reflecting differences in daily or hourly weather conditions. With this in mind, some of the limitations imposed by the CES model can be overcome. By being able to account for differences between normal days and event days through the specification of categorical variables, we can also go part way toward allowing the elasticities of substitution to differ on the basis of the peak to off-peak price ratio. We outline such an alternative next.

10.4 The CES Model

To develop this model, we begin by assuming that the customer's separable, sub-utility function, U, for the consumption of electricity is given by:

Eq. D-1: $U = A\left[\delta K_p^{-\rho} + (1-\delta)K_o^{-\rho}\right]^{-1/\rho}$

where K_p and K_o are consumption of peak and off-peak electricity, respectively, and A, δ, and ρ are parameters to be estimated. Furthermore, assume that the peak and off-peak prices for electricity are r_p and r_o, respectively, and that the customer's electricity budget is B. The customer's decision problem is to maximize utility from the consumption of peak and off-peak electricity subject to the budget allocated to electricity consumption. That is, it is to maximize:

Eq. D-2: $L = A\left[\delta K_p^{-\rho} + (1-\delta)K_o^{-\rho}\right]^{-1/\rho} + \lambda[B - r_p K_p - r_o K_o]$

Chang and Wainwright (2005) derive convenient expressions for the partial derivatives of utility with respect to the two goods—in our case, peak and off-peak electricity. Using their results, the first-order necessary conditions for a maximum are:

Eq. D-3: $\dfrac{\partial U}{\partial K_p} = \dfrac{\delta}{A^{\rho}}\left(\dfrac{U}{K_p}\right)^{1+\rho} - r_p = 0$

Eq. D-4: $\dfrac{\partial U}{\partial K_o} = \dfrac{(1-\delta)}{A^{\rho}}\left(\dfrac{U}{K_o}\right)^{1+\rho} - r_o = 0$;

Eq. D-5: $\dfrac{\partial L}{\partial \lambda} = \left[B - r_p K_p - r_p K_p\right] = 0$.

Then, forming the ratio of the first-order conditions (Equations D-3 and D-4), we have:

Eq. D-6: $$\frac{\dfrac{(1-\delta)}{A^{\rho}}\left(\dfrac{U}{K_o}\right)^{1+\rho}}{\dfrac{\delta}{A^{\rho}}\left(\dfrac{U}{K_p}\right)^{1+\rho}} = \frac{r_o}{r_p}.$$

After some reorganization, Equation D-6 reduces to:

Eq. D-7: $$\frac{(1-\delta)\left(K_p\right)^{1+\rho}}{\delta\left(K_o\right)^{1+\rho}} = \frac{r_o}{r_p},\ \frac{\left(K_p\right)^{1+\rho}}{\left(K_o\right)^{1+\rho}} = \frac{\delta r_o}{(1-\delta) r_p}$$

10.4.1 Identifying the Elasticity of Substitution

By raising both sides of Equation D-7 to the $1/(1+\rho)$ power, we solve for the ratio of peak to off-peak electricity consumption as a function of the ratio of off-peak to peak prices:

Eq. D-8: $$\frac{\left(K_p\right)}{\left(K_o\right)} = \left(\frac{\delta}{1-\delta}\right)^{\frac{1}{1+\rho}}\left(\frac{r_o}{r_p}\right)^{\frac{1}{1+\rho}}$$

Chang and Wainwright (2005) show that $\sigma = 1/(1+\rho)$ is the elasticity of substitution—in our case, the percentage change in the ratio of peak to off-peak consumption resulting from a 1% change in the off-peak to peak price of electricity. By taking the logarithm of this equation and substituting σ for $1/(1+\rho)$, we can confirm this interpretation of σ:

Eq. D-9: $$\ln\frac{\left(K_p\right)}{\left(K_o\right)} = \sigma \ln\left(\frac{\delta}{1-\delta}\right) + \sigma \ln\left(\frac{r_o}{r_p}\right)$$ [64]

[64] It is well-known that the CES model is self-dual in the sense that an indirect utility function that could be derived from the maximization problem in Eq. D-2 has the same algebraic form as the direct utility function in Eq. D-1. Therefore, if one began this derivation with a dual indirect CES utility function, Eq. D-9 still depicts the relationship between the ratio of peak to off-peak electricity use and the inverse price ratio.

10.4.2 The Estimating Equation

By letting $\gamma = \sigma \ln\left(\frac{\delta}{1-\delta}\right)$, we can simplify the expression to:

Eq. D-10: $\ln\frac{(K_p)}{(K_o)} = \gamma + \sigma \ln\left(\frac{r_o}{r_p}\right)$

Although we cannot obtain an estimate of δ, we are able to add an error term, u, to this equation and obtain an estimate of the elasticity of substitution, σ, by ordinary least squares. Although we suppress the subscript $i = 1,\ldots,m$ to indicate the individual customer, let us assume that we have a pooled data set that consists of a time series of cross sections and that there are $t = 1,\ldots,n$ daily observations for each customer, $i = 1,\ldots,m$.

To simplify the algebra, let us assume a slightly simpler model than the one discussed in the text. That is, let us assume that there are only two rate treatments in this pilot. Customers in the first treatment face a real-time pricing (RTP) rate in which the 24-hourly prices are announced a day ahead. However, although customers in the second rate treatment generally face the RTP prices, there are a certain number of critical peak event days that can be called during the summer months. On these event days, Treatment 2 customers pay very high CPP prices during the peak hours of the day (call the rate *RTP-CPP*). Then, we can begin with the model:

Eq. D-11: $\ln\frac{(K_{pti})}{(K_{oti})} = \gamma + \sigma \ln\left(\frac{r_{oti}}{r_{pti}}\right) + u_{ti}$

for every customer regardless of the rate treatment. However, as demonstrated previously, we can account for the effect of several different exogenous conditions and customer and premise circumstances by enriching the model with additional variables. [65] At the end of this discussion, we illustrate the way in which these estimates of the elasticities of substitution can also be used to estimate t constant utility, and cross-price elasticities of demand for peak and off-peak electricity usage. For this demonstration, it is sufficient to specify a model that contains only a subset of the effects included in the model in the text.

For this model, we again account for the effects of different weather conditions on the demand for peak and off-peak electricity by augmenting Eq. D-11 with two weather variables. The first, CD_t, measures cooling degrees, and it enters as an intercept shifter—controlling for differences

[65] In this model, one could also include dummy variables to control for the effects of groups of customers or customer-specific characteristics. However, if we also exploit the pooled nature of the data and estimate a fixed-effects model, the only way to identify the effects of explanatory variables that are time-invariant for each customer is to assume that these customer or premise characteristics (for example, air conditioning) affect usage primarily through their interaction with treatments and/or factors (for example, weather) that do vary over time. Therefore, these effects can be captured by the coefficient associated with variables created by constructing interaction terms between these types of variables with those that represent the various treatments. Although we do not specify this as a fixed-effects model, the model specification is consistent with this way of capturing the effects of these customer-specific characteristics. If we were to specify a fixed-effects model, we would add a term μ_i to the models in Equation D-11 and use a fixed-effects estimator to estimate the model.

in peak to off-peak usage as temperature changes. Variable H_t takes on the binary values of unity for hot days and zero otherwise, where a *hot day* is defined as one in which a heat index is above 85°F (29°C).[66] We can also specify a binary variable CPP_t that takes on a value of unity for customers on the RTP-CPP rate and zero otherwise. Finally, assume that the binary variable (AC_i) takes on a value of unity if the customer has central air conditioning and zero otherwise. By including these two weather-related variables—the rate treatment variable and the variable for central air in the model—the estimating equation takes the following form:[67]

Eq. D-12: $$\ln\frac{(K_{pti})}{(K_{oti})} = \gamma + cd(CD_t) + [\sigma + a(AC_i) + h(H_t) + c(CPP_t)]\ln\left(\frac{r_{oti}}{r_{pti}}\right) + u_{ti}$$

To estimate this model, there are now six variables: one dependent variable, five independent variables, and a constant term γ.[68]

The parameters associated with these variables are evident from Equation D-12. Accordingly, because the variable (CD_t) measures the cooling degrees, one might expect that as this variable increases *ceteris paribus*, there would be an increase in the ratio of the hourly peak to off-peak electricity use. Therefore, one might expect the estimated value of the coefficient on this variable to be positive, $cd > 0$.

10.4.3 The Estimated Elasticities of Substitution

To interpret the other variables, it is essential to recognize that they affect the elasticity of substitution in the following way. Define the effective elasticity of substitution as σ^E. Accordingly, we have:

Eq. D-13: $$\hat{\sigma}^E = \hat{\sigma} + \hat{h}H_t + \hat{c}EV_t + \hat{a}(AC_i)$$

where the "hat" (^) indicates an estimated coefficient. Because customers that rely on air conditioning to keep cool on hot summer days, we might expect those customers to be less willing to shift load off peak, and the elasticity of substitution would be lower than for customers with no air conditioning—therefore, $a < 0$. If one expects that the customer's willingness to shift load from peak to off-peak falls, one would expect the coefficient on the variable for H_t to be negative as well: $h < 0$. Further, if one would expect a customer's willingness to shift load from peak to off-peak during event days (over and above the response resulting from higher prices), one would expect that the coefficient on the event day variable would be $c > 0$. This effect is in the opposite direction from that for the hot day. If both circumstances occur on the same day, the anticipated effects may partially offset one another. In this model, there are now eight separate elasticities of substitution (one for each combination of values for the three dummy variables); similar to the discussion in the text, any of the separate estimates of the elasticity of substitution

[66] The specific construction of these two weather variables is primarily for illustration purposes and follows Boisvert et al. (2007).

[67] There could also be some alternative specifications in which CD_t could affect the elasticity of substitution or where H_t would enter as an intercept shifter and affect the ratio of peak to off-peak electricity consumption directly as well as indirectly through the elasticity of substitution.

[68] For completeness, these variables are listed in the discussion of this example in the text.

can be derived from Equation D-13. The coefficients that appear in a particular estimate depend on which of the dummy variables take on values of unity and which take on values of zero.

10.4.4 Estimating the Compensated Own- and Cross-Price Elasticities

From this CES model, we can also estimate the compensated own-price elasticities of peak and off-peak electricity usage. Using the same example from above, we first solve Equation D-14:

Eq. D-14: $$\ln\frac{\left(\hat{K}_{pti}\right)}{\left(\hat{K}_{oti}\right)} = \hat{\gamma} + \hat{cd}(CD_t) + \left[\hat{\sigma} + \hat{h}(H_t) + \hat{a}(AC_i) + \hat{c}(CPP_t)\right]\ln\left(\frac{r_{oti}}{r_{pti}}\right)$$

for the peak to off-peak ratio of electricity use. The result is:

Eq. D-15: $$\frac{\left(\hat{K}_{pti}\right)}{\left(\hat{K}_{oti}\right)} = e^{\hat{\gamma}+\hat{cd}(CD_t)}\left(\frac{r_{oti}}{r_{pti}}\right)^{\hat{\sigma}+\hat{h}(H_t)+\hat{a}(AC_i)+\hat{c}(CPP_t)}$$

Furthermore, we can solve for the K_{oti}:

Eq. D-16: $$\left(\hat{K}_{oti}\right) = \frac{\left(\hat{K}_{pti}\right)}{\left(e^{\hat{\gamma}+\hat{cd}(CD_t)}\left(\frac{r_{oti}}{r_{pti}}\right)^{\hat{\sigma}+\hat{h}(H_t)+\hat{a}(AC_i)+\hat{c}(CPP_t)}\right)}$$

We also know what the actual total expenditures for electricity are on any given day, B_{ti}; therefore:

Eq. D-17: $$B_{ti} = r_{pti}K_{pti} + r_{oti}K_{oti}$$

Substituting Equation D-16 into Equation D-17, we have:

Eq. D-18: $$B_{ti} = r_{pti}\left(\hat{K}_{pti}\right) + r_{oti}\frac{\left(\hat{K}_{pti}\right)}{\left(e^{\hat{\gamma}+\hat{cd}(CD_t)}\left(\frac{r_{oti}}{r_{pti}}\right)^{\hat{\sigma}+\hat{h}(H_t)+\hat{a}(AC_i)+\hat{c}(CPP_t)}\right)}$$

Now, given that we know the prices, we can solve for K_{pti}:

Eq. D-19: $B_{ti} = r_{pti}\left(\hat{K}_{pti}\right) + r_{oti}\dfrac{\left(\hat{K}_{pti}\right)}{\left(e^{\hat{\gamma}+\hat{c}d(CD_t)}\left(\dfrac{r_{oti}}{r_{pti}}\right)^{\hat{\sigma}+\hat{h}(H_t)+\hat{a}(AC_i)+\hat{c}(CPP_t)}\right)}$

We can then substitute this result into Equation D-17 to calculate K_{oti}. That is:

Eq. D-20: $B_{ti} = r_{pti}\left(\hat{K}_{pti}\right) + r_{oti}\dfrac{\left(\hat{K}_{pti}\right)}{\left(e^{\hat{\gamma}+\hat{c}d(CD_t)}\left(\dfrac{r_{oti}}{r_{pti}}\right)^{\hat{\sigma}+\hat{h}(H_t)+\hat{a}(AC_i)+\hat{c}(CPP_t)}\right)}$

The estimated expenditure shares would follow immediately.

Based on these estimated expenditure shares, we can now calculate these "own-Allen" partial elasticities of substitution[69] from the expenditure shares and the "cross-Allen: partial elasticity of substitution. In the case of two goods, it these are equivalent to the elasticity of substitution given in Equation D-13. As Berndt (1991) demonstrates, when own-Allen elasticities are multiplied by expenditure shares, one has expressions for the compensated own-price elasticities of demand—the percentage change in the demand for peak or off-peak electricity resulting from percentage changes in their own prices that will leave a customer's utility unchanged. The elasticities are $E_{pp} = w_p\ \sigma_{pp}$ and $E_{oo} = w_o\ \sigma_{oo}$, respectively.

Furthermore, compensated cross-price elasticities (the percentage change in peak [off-peak] demand resulting from a percentage change in the off-peak [peak] price) are similarly given by: $E_{po} = ES_o\ \sigma_{po}$ and $E_{op} = ES_p\ \sigma_{op}$, respectively. Because these compensated price elasticities must satisfy the adding of conditions to ensure homogeneity of demand, we also know that $ES_p\ \sigma_{pp} + ES_o\ \sigma_{po} = 0$ and $ES_p\ \sigma_{op} + ES_o\ \sigma_{oo} = 0$. Therefore, we can solve for $\sigma_{pp} = -(ES_o/ES_p)\ \sigma_{po}$ and $\sigma_{oo} = -(ES_p/ES_o)\ \sigma_{po}$. By calculating the own-Allen elasticities in this way, one can also ensure the internal consistency of the empirical results.

10.5 The Generalized Leontief Model

As suggested previously, the use of the generalized Leontief (GL) regression models to estimate these changes in load (measured by elasticities of substitution) allows us to control for each customer's typical load profile and for the effects of weather. These customer-specific

[69] The Allen partial elasticity measures the degree to which the demand for factor *j* changes as the price of factor *i* changes (Allen 1938, pp. 508–09). If $\sigma_{ij} > 0$ and the price of factor *i* increases, the use of factor *j* increases—taking part in the replacement of factor *i* in production. The two factors are said to be *competitive*. If $\sigma_{ij} < 0$, the two factors are *complements*, as the price of one rises, the demands for both fall.

regressions, based on the GL demand model, provide estimates of the elasticities of substitution that differ by day.[70]

This is an attractive feature of the model, and it is exploited by estimating customer-specific regressions—particularly for the subset of customers identified as *responders*. The daily estimates of the elasticities of substitution can then be used in a second-stage, meta-analysis of the sub-set of customers identified as *responders*. The strategy is to pool the data for all *responders*. Before discussing this meta-analysis, it is essential to develop the GL model; this is accomplished in a straightforward fashion using a two-commodity specification of the GL model.[71]

10.5.1 The Two-Commodity Model for Peak and Off-Peak Electricity Demand

If we define P_p and P_o as the prices of peak and off-peak electricity, respectively, and ES_p and ES_o as the shares of electricity expenditure in peak and off-peak periods, respectively, the two-commodity homothetic GL indirect utility function and budget shares for electricity can be written as follows:

Eq. D-21: $$V = \frac{Y}{\gamma_{pp}P_p + 2\gamma_{po}\sqrt{P_p P_o} + \gamma_{oo}P_o}$$

Eq. D-22: $$ES_p = \frac{\gamma_{pp}P_p + \gamma_{po}\sqrt{P_p P_o}}{\gamma_{pp}P_p + 2\gamma_{po}\sqrt{P_p P_o} + \gamma_{oo}P_o}$$

---

70 This section is based largely on the development of the GL model by R. Boisvert that is found in the evaluation of the ComEd CAP program (EPRI 1023644, Appendix A).

71 This issue is discussed by Boisvert et al. (2007). Their discussion is repeated here for completeness. Effectively, they begin by suggesting that there were several other second-order flexible forms that might have been used in this empirical specification. One such commonly used flexible form, the translog (TL) model (Boisvert 1982 and Chambers 1988), would have avoided estimating any equations that are nonlinear in the parameters. The TL model relies on estimating a set of electricity cost-share equations that are linear in the model parameters and does not require observations on the electricity aggregate. Although this TL form was particularly easy to estimate, it was not pursued. Caves and Christensen (1980 a, b) point out that the TL model does not perform well when substitution elasticities are likely to be small or when there are likely to be small shares or large relative differences among shares. Patrick and Wolak (2001) found this to be problematic in an application of customer demand for electricity under real-time pricing.

Patrick and Wolak (2001) argue that the GL model is superior to the TL model because it has a fixed-coefficient Leontief technology as a limiting case, and therefore, it can reflect rather modest substitution possibilities. They also note that if one imposes global concavity, all inputs in the GL model must be substitutes. To circumvent these difficulties, Patrick and Wolak (2001) and Taylor *et al.* (2005) employ a generalized McFadden (GM) cost function that is "…second-order flexible, yet suited to capture small positive and negative elasticities of substitution between electricity demands across load periods within a day" (Patrick and Wolak 2001, p. 27). Since our empirical model specifies only two demand periods, any model in which global concavity is either assumed, or imposed, requires that the inputs be substitutes. Therefore, the primary issue that led to their selection of the GM model is of no concern in our application of the GL model.

Eq. D-23: $ES_o \quad = \quad \dfrac{\gamma_{oo} P_o + \gamma_{po}\sqrt{P_p P_o}}{\gamma_{pp} P_p + 2\gamma_{po}\sqrt{P_p P_o} + \gamma_{oo} P_o}$

The budget share equations are homogeneous of degree zero in γ_{pp}, γ_{po}, and γ_{oo}.[72] Therefore, without loss of generality, we can adopt the normalization of $\gamma_{pp} + 2\gamma_{po} + \gamma_{oo} = 1$.
For the two-commodity homothetic GL model, Caves and Christensen (1980a) show that the elasticity of substitution between peak and off-peak electricity can be written as:

Eq. D-24: $\sigma_{po} \quad = \quad \dfrac{\gamma_{po}\sqrt{P_p P_o}}{2ES_p\left(1 - ES_p\right)\left(\gamma_{pp} P_p + 2\gamma_{po}\sqrt{P_p P_o} + \gamma_{oo} P_o\right)}.$

By examining Eq. D-24, it is clear that for any given set of parameters (either assumed or estimated from data), the elasticity of substitution is a function of both the estimated parameters of the function and the prices. Furthermore, σ, ES_p, and ES_o can all be nonnegative only if γ_{po} is nonnegative.[73]

There is no restriction, however, on the signs of γ_{pp} and γ_{oo}.[74]

In the extreme case in which $\gamma_{po} = 0$, the elasticity of substitution is zero. The null hypothesis that the elasticity of substitution between peak and off-peak electricity consumption is zero is therefore conveniently tested using the GL demand model.

10.5.2 The Estimating Equations

---

[72] *Homogeneity of degree zero* means that budget shares do not change if all prices and expenditures change in the same proportion.

[73] This is sufficient for the indirect utility function to satisfy the monotonicity requirement; that is, $\partial V/\partial p_i < 0$ for all i.

74 We must also test that the quasi-convexity requirement on V is met at each data point. This is equivalent to the requirement that the matrix of Allen partial elasticities of substitution be negative semi-definite (Berndt 1991). The cross-Allen partial elasticities of substitution are symmetric: $\sigma_{po} = \sigma_{op}$. To perform the test for quasi-convexity, however, we must also calculate the own-Allen elasticities of substitution, σ_{pp} and σ_{oo}, so that we have the complete matrix of Allen elasticities.

To reiterate, for the two-goods case, one can calculate these own-Allen partial elasticities from the expenditure shares and the cross-Allen partial elasticity of substitution. One can first recall from Berndt (1991) that when these own-Allen elasticities are multiplied by expenditure shares, one has expressions for the compensated own-price elasticities of demand—the percentage change in the demand for peak or off-peak electricity resulting from percentage changes in their own prices that will leave a customer's utility unchanged; the elasticities are $E_{pp} = w_p\,\sigma_{pp}$ and $E_{oo} = w_o\,\sigma_{oo}$, respectively. Furthermore, compensated cross-price elasticities (the percentage change in peak [off-peak] demand resulting from a percentage change in the off-peak [peak] price) are similarly given by $E_{po} = ES_o\,\sigma_{po}$ and $E_{op} = ES_p\,\sigma_{op}$, respectively. Because these compensated price elasticities must satisfy the adding of conditions to ensure homogeneity of demand, we also know that $ES_p\,\sigma_{pp} + ES_o\,\sigma_{po} = 0$ and $ES_p\,\sigma_{op} + ES_o\,\sigma_{oo} = 0$. Therefore, we can solve for $\sigma_{pp} = -(ES_o/ES_p)\,\sigma_{po}$ and $\sigma_{oo} = -(ES_p/ES_o)\,\sigma_{po}$. By calculating the own-Allen elasticities in this way, one can also ensure the internal consistency of the empirical results.

To estimate the elasticities of substitution in Eq. D-24, we must have estimates of the parameters of the indirect utility function in Eq. D-21. It cannot be estimated directly because we have no data for customers' utility levels *V*. Because we can derive the electricity expenditure share equations from Eq. D-21, we can obtain estimates of the parameters of *V* by estimating the share Eq. D-22 and Eq. D-23. However, a strategy for doing this is not completely straightforward. To estimate the share equations directly would require a nonlinear systems estimator that applies cross-equation constraints on every parameter. It is necessary to impose the normalization on the parameters by adding a constraint in which $\gamma_{pp} + 2\gamma_{po} + \gamma_{oo} = 1$.

As an alternative, we can simplify the estimation by first forming an equation for the ratio of expenditure shares. Because the denominators of these share equations are identical, we are left with a single equation that is simpler than either Eq. D-22 or Eq. D-23:

Eq. D-25: $$\frac{ES_p}{ES_o} = \frac{\gamma_{pp}P_p + \gamma_{po}\sqrt{P_p P_o}}{\gamma_{oo}P_o + \gamma_{po}\sqrt{P_p P_o}}$$

This is still an equation that is nonlinear in the parameters, but the estimation is reduced to that of a single equation in which only symmetry need be imposed on the coefficient γ_{po}. It is still necessary to impose the normalization on the parameters by adding a constraint in which $\gamma_{pp} + 2\gamma_{po} + \gamma_{oo} = 1$. Furthermore, past experience (for example, Boisvert et al. 2007, Braithwait 2000, and Caves and Christensen 1984 a, b) suggests that the estimation is facilitated by transforming the equation into logarithms, as follows:

Eq. D-26: $$\ln\left(ES_p / ES_o\right) = \ln\left[\gamma_{pp}P_p + \gamma_{po}\sqrt{P_p P_o}\right] - \ln\left[\gamma_{oo}P_o + \gamma_{po}\sqrt{P_p P_o}\right] + u$$

where *u* is a stochastic error term.

10.5.3 An Empirical Specification

To conduct the estimation, we must first use the customers' hourly data on electricity prices and usage to create observations of peak and off-peak expenditure shares and average prices (perhaps weighted by hourly usage). This will create one observation per customer per day. By letting the observations for the t^{th} day be denoted by the subscript *t*, we have for each customer in each treatment group the following equation:

Eq. D-27: $$\ln\left(ES_{pt} / ES_{ot}\right) = \ln\left[\gamma_{pp}P_{pt} + \gamma_{po}\sqrt{P_{pt} P_{ot}}\right] - \ln\left[\gamma_{oo}P_{ot} + \gamma_{po}\sqrt{P_{pt} P_{ot}}\right] + u_t$$

To normalize the coefficients, this equation would be estimated subject to the following constraints—the latter of which is required by Young's Theorem (which states that under certain conditions second derivatives are symmetric):

Eq. D-27a: $\gamma_{pp} + 2\gamma_{po} + \gamma_{oo} = 1$, $\gamma_{op} = \gamma_{po}$

Because of the large numbers of customers in each treatment and of daily observations per customer, we estimate the model for each customer in each treatment. In this way, customer-level effects are reflected in the daily estimates of the elasticities of substitution. Tests of the differences in the elasticities of substitution by treatment or by customer characteristic will be conducted in a second regression analysis. In this analysis, the daily estimates of the elasticities of substitution by customer will be pooled across customers and treatments. These pooled estimates will be regressed against peak to off-peak price ratios, variables controlling for individual customer characteristics, and dummy variables to control for the rate treatment effects. This strategy is similar to that employed by Boisvert et al. (2007) and Taylor and Schwarz (1990). The exact specification of this regression will depend on the nature of customer-specific data available through the survey responses or other sources of data. It is possible and appropriate to reflect differences in daily weather conditions in the regressions for individual customers. By including two additional variables, Boisvert et al. (2007) control for the effects of weather. The first, CD_t, measures cooling degrees, and it enters as an intercept shifter—controlling for differences in peak to off-peak usage as temperature changes. Variable H_t takes on the binary values of unity for hot days and zero otherwise, where a *hot day* is defined as one in which a heat index is above 85°F (29°C). Assuming that two weather-related variables such as these are included, the estimating equations become:

Eq. D-28:
$$\begin{aligned}&\ln\left(ES_{pt}/ES_{ot}\right) = cd(CD_t) + \ln\left[h_p H_t + \gamma_{pp} P_{pt} + \gamma_{po}\sqrt{P_{pt}P_{ot}}\right]\\ &- \ln\left[h_o H_t + \gamma_{oo} P_{ot} + \gamma_{po}\sqrt{P_{pt}P_{ot}} + \right] + u_t\end{aligned}$$

We also need the following:

Eq. D-28a: $\gamma_{pp} + 2\gamma_{po} + \gamma_{oo} = 1$
Eq. D-28b: $h_p = h_o$
Eq. D-28c: $\gamma_{op} = \gamma_{po}$

10.5.4 Estimating the Daily Elasticities of Substitution

Once Equation D-28 is estimated, the parameter estimates are used to calculate the daily elasticities of substitution between peak and off-peak electricity use. As is evident from Equation D-24, we must first calculate predicted values for the peak and off-peak expenditure shares. In the standard formulation of the GL model, one could simply substitute the estimated parameters into Equations D-22 and D-23. However, once the weather variables have been introduced into the model, one must calculate the predicted value for ES_{pt} directly from Equation D-28 in the following way.

We can substitute the estimated parameters (denoted by "^") directly into Equation D-28:

Eq. D-29:
$$\begin{aligned}&\ln(\hat{ES}_{pt}/\hat{ES}_{ot}) = \hat{cd}\,CD_t + \ln[\hat{h}_p H_t + \hat{\gamma}_{pp} P_{pt} + \hat{\gamma}_{po}\sqrt{P_{pt}P_{ot}}]\\ &- \ln[\hat{h}_o H_t + \hat{\gamma}_{oo} P_{ot} + \hat{\gamma}_{po}\sqrt{P_{pt}P_{ot}}]\end{aligned}$$

Taking the anti-log of Equation D-29, we have:

Eq. D-30: $\hat{ES}_{pt} / \hat{ES}_{ot} = [e^{\hat{cd}\, CD_t}] \dfrac{[\hat{h}_p H_t + \hat{\gamma}_{pp} P_{pt} + \hat{\gamma}_{po} \sqrt{P_{pt} P_{ot}}\,]}{[\hat{h}_o H_t + \hat{\gamma}_{oo} P_{ot} + \hat{\gamma}_{po} \sqrt{P_{pt} P_{ot}}\,]}$

Because the two expenditure shares sum to unity,

Eq. D-31: $\hat{ES}_{pt} + \hat{ES}_{ot} = 1$

we can solve Eq. D-31 for the off-peak expenditure share and substitute the result into Eq. D-30. After some rearranging, we have:

Eq. D-32: $\hat{ES}_{pt} = [e^{\hat{cd}\, CD_t}] \dfrac{[\hat{h}_p H_t + \hat{\gamma}_{pp} P_{pt} + \hat{\gamma}_{po} \sqrt{P_{pt} P_{ot}}\,]}{[\hat{h}_o H_t + \hat{\gamma}_{oo} P_{ot} + \hat{\gamma}_{po} \sqrt{P_{pt} P_{ot}}\,]} (1 - \hat{ES}_{pt})$

Solving this equation for the predicted peak expenditure share, we have:

Eq. D-33: $$\hat{ES}_{pt} = \frac{[e^{\hat{cd}\, CD_t}] \dfrac{[\hat{h}_p H_t + \hat{\gamma}_{pp} P_{pt} + \hat{\gamma}_{po} \sqrt{P_{pt} P_{ot}}\,]}{[\hat{h}_o H_t + \hat{\gamma}_{oo} P_{ot} + \hat{\gamma}_{po} \sqrt{P_{pt} P_{ot}}\,]}}{[1 + e^{\hat{cd}\, CD_t}] \dfrac{[\hat{h}_p H_t + \hat{\gamma}_{pp} P_{pt} + \hat{\gamma}_{po} \sqrt{P_{pt} P_{ot}}\,]}{[\hat{h}_o H_t + \hat{\gamma}_{oo} P_{ot} + \hat{\gamma}_{po} \sqrt{P_{pt} P_{ot}}\,]}}$$

In the standard GL model (without the weather variables), we know that the underlying utility function is globally quasi-convex if all of the γ parameters are strictly positive. In this case, we are guaranteed that this peak expenditure share is between zero and unity. By including the weather variables, it is clear that the peak expenditure share—and therefore the off-peak expenditure share—will differ by hot and cool days and by the number of cooling degrees, depending on their values for these days. Furthermore, we are guaranteed that this peak expenditure share is between zero and unity if the terms $h_p H_t$ and $h_o H_t$ and parameters are positive as well.

If these two terms are negative, their absolute values must be smaller than the corresponding positive expressions in Eq. D-33 in the numerators and denominators of that equation to guarantee that the shares are positive and add to unity. Because of the complexity added to the model by the inclusion of these weather variables, we must evaluate the principal minors of the matrix of Allen[75] partial elasticities of substitution to determine whether the underlying utility

[75] Again recall that the Allen partial elasticity measures the degree to which the demand for good *j* changes as the price of good *i* changes (Allen 1938, pp. 508–09). If $\sigma_{ij} > 0$ and the price of good *i* increases, the use of good *j* increases—taking part in the replacement of factor *i* in production. The two factors are said to be *competitive*. If, $\sigma_{ij} < 0$, the two factors are *complements*: as the price of one rises, the demands for both fall.

function is quasi-convex at every data point. In all cases, it is necessary for $\gamma_{po} > 0$, but this need not necessarily be the case for γ_{pp} or γ_{oo}.

Finally, these estimated relationships can be substituted into the expressions for the daily Allen elasticity of substitution between peak and off-peak electricity consumption:

Eq. D-34: $$\hat{\sigma}_{pot} = \frac{\hat{\gamma}_{po}\sqrt{P_{pt}P_{ot}}}{2\hat{ES}_{pt}(1-\hat{ES}_{pt})(\hat{\gamma}_{pp}P_{pt} + 2\hat{\gamma}_{po}\sqrt{P_{pt}P_{ot}} + \hat{\gamma}_{oo}P_{ot})}.$$

From Eq. D-33, it is evident that the two weather variables affect the size of the peak expenditure share. In turn, the size of the elasticity of substitution is also a function of this share because of the term:

Eq. D-35: $$\frac{1}{\hat{ES}_{pt}(1-\hat{ES}_{pt})}$$

This term decreases [as does $(\hat{\sigma}_{po})$] as $\hat{ES}_{pt}$ increases for $0 < \hat{ES}_{pt} < 0.5$. The term reaches a minimum at $\hat{ES}_{pt} = 0.5$ and then increases [as does $(\hat{\sigma}_{po})$] for $0.5 < \hat{ES}_{pt} < 1.0$.
It follows immediately that one can also calculate internally consistent estimates of the own Allen elasticities of substitution that are needed to check the quasi-convexity requirement on *V* at each data point:

Eq. D-36: $$\hat{\sigma}_{ppt} = \frac{\hat{ES}_{ot}}{\hat{ES}_{pt}}\hat{\sigma}_{pot}$$

Eq. D-37: $$\hat{\sigma}_{oot} = \frac{\hat{ES}_{pt}}{\hat{ES}_{ot}}\hat{\sigma}_{pot}$$

To check the quasi-convexity requirement on *V* at each data point, we need to check that (Berndt 1991):

Eq. D-38: $\hat{\sigma}_{ppt} \leq 0$; and $\hat{\sigma}_{oot} \leq 0$; and $(\hat{\sigma}_{ppt})(\hat{\sigma}_{oot}) - (\hat{\sigma}_{pot})^2 \geq 0$

10.6 The Meta-Analyses

As suggested above, the use of the GL regression models to estimate these changes in load (measured by elasticities of substitution) allows for us to control for each customer's typical load profile and for the effects of weather. These customer-specific regressions, based on the GL demand model, provide estimates of the elasticities of substitution that differ by day.

This is an attractive feature of the model, and it is exploited by estimating customer-specific regressions. The daily estimates of the elasticities of substitution can then be used in a second-stage meta-analysis—the strategy is to pool the data for all customers. Using this pooled data set, a regression is estimated that has these daily elasticities of substitution as left-hand side variables; the independent variables include those designed to explain these differences in the daily elasticities of substitution.[76] Because the effects of weather are already accounted for in the GL specification, the other factors would include premise circumstances (some would be available for all customers), enabling technologies, and customer demographics.[77] An important part of this analysis is to include the daily ratio of peak to off-peak prices as an explanatory variable. By including this variable, we are able to determine whether—after controlling for all other factors—the elasticities of substitution are in fact affected by the price level. That is, are customers more price responsive as the peak to off-peak price ratio rises? These variables are specified in the second regression model as intercept and/or slope shifters to account for the interaction among the characteristics. Because differences in weather are already accounted for in the model, we can ignore it for the specification of this equation. The equation for the meta-analysis, written in its most general form, is:[78]

Eq. D-39:

$$\sigma_{Ht} = \alpha + \beta_1(r_{pt}/r_{ot}) + \sum_{j-1}\beta_j(Rate\ Treatment_{jH}) + \sum_{k-1}\beta_k(Device\ or\ Education\ Treatment_{kH}) + \sum_{m-1}\beta_k(Customer\ Characteristic_{mH}) + \sum_{n-1}\beta_n(PremiseCharacteristic_{nH}) + \varepsilon_t$$

where σ_{Ht} is the elasticity of substitution for customer *H* on day *t* and (r_{pt}/r_{ot}) is the ratio of peak to off-peak price on day *t*. There are assumed to be *J* rate treatments, *K* device or education treatments, *M* customer characteristics, and *N* premise characteristics. The effects of these factors on the individual customers' elasticities of substitution are captured by individual (0, 1) variables

---

76 This strategy for analyzing these types of results is similar to that used by Taylor and Schwarz (1990); they estimated a GL model of household demand response. They pooled estimates of the parameters from the GL model for each cross section in their data. Using each set of estimated parameters as the dependent variable in a regression model, they determined the extent to which the parameter estimates are affected by weather and a customer's prior experience in a TOU program.

77 The ability to include customer demographics depends on the availability of data from survey respondents as well as the response rates for customers.

78 A similar strategy was used by Boisvert et al. (2007) in a study of large industrial and commercial customers. The specification of that model was as follows:

$$\sigma_{tf} = \alpha + \beta_1[p_{pt,f}/p_{0t,f}] + \beta_2 \text{Man}^*[\%\text{MaxD}_{t,f}] + \beta_3 \text{Man}^*[p_{pt,f}/p_{0t,f}] + \beta_4 \text{Gov/Ed}^*[\%\text{MaxD}_{t,f}] + \beta_5 \text{Gov/Ed}^*[p_{pt,f}/p_{0t,f}] + \beta_6 \text{PW}^*[\%\text{MaxD}_{t,f}] + \beta_7 \text{PW}^*[p_{pt,f}/p_{0t,f}] + \beta_8 \text{C\&R}^*[\%\text{MaxD}_{t,f}] + \beta_9 \text{C\&R}^*[p_{pt,f}/p_{0t,f}] + u_{tf}$$

where $[p_{pt,f}/p_{0t,f}]$ = ratio of peak to off-peak electricity prices on day *t* for firm *f*; Man = 1 for manufacturing firms, 0 otherwise; Govt/Ed = 1 for government/education firms, 0 otherwise; PW = 1 for public works firms, 0 otherwise; C&R = 1 for commercial and retail firms, 0 otherwise; $\%MaxD_{t,f}$ = % peak use on day *t* as a percentage of maximum demand (of the summer) for firm *f*; and utf = a random error term.

for each of the factors. To avoid singularities in the regression equation, one of the categories in each group is dropped, so that the coefficient α is the intercept for the reference customer whose factors correspond to those that are dropped from the regression.

The regression specified in Eq. D-39 accounts for many factors that could affect a customer's willingness to shift load, as measured by the elasticity of substitution. However, in its specification, one of the explanatory variables is the ratio of peak to off-peak prices. These price ratios will differ daily if the rate treatment is an RTP; in this case, the specification in Eq. D-39 may be appropriate. However, these price ratios will differ only on an event day if the rate treatment is a CPP but the background rate is a flat rate or a TOU. Therefore, in these cases there is no reason to test for the effect of daily price ratios on the elasticities of substitution. In these cases we may, however, wish to test some important hypotheses about the effects of rate, device, and education treatments as well as the customer or premise characteristics on the average elasticities of substitution for each customer. In this case, we can estimate an alternative model. This second model involves pooling data for the average estimated elasticities of substitution across customers and regressing the average elasticities of substitution on individual rate treatment, device, and educational treatments as well as customer and premise characteristics. In general form, the model would be:

Eq. D-40:

$$\sigma_{HA} = \alpha + \sum_{j-1} \beta_j \left(Rate\ Treatment_{jH}\right) + \sum_{k-1} \beta_k (Device\ or\ Education\ Treatment_{kH}) + \sum_{m-1} \beta_m (Customer\ Characteristic_{mH}) + \sum_{n-1} \beta_n (PremiseCharacteristic_{nH}) + \varepsilon$$

where σ_{HA} is the average daily elasticity of substitution for customer *H.* Again, there are assumed to be *J* rate treatments, *K* device or education treatments, *M* customer characteristics, and *N* premise characteristics. The effects of these factors on the individual customers' average elasticities of substitution are captured by individual (0, 1) variables for each of the factors. To avoid singularities in the regression equation, one of the categories in each group is dropped, so that the coefficient α is the intercept for the reference customer whose factors correspond to those that are dropped from the regression.