Title

Winners are not keepers: Characterizing household engagement, gains, and energy patterns in demand response using machine learning in the United States

Authors

Todd-Blick, Annika
Spurlock, C Anna
Jin, Ling
et al.

Publication Date

2020-12-01

DOI

10.1016/j.erss.2020.101595

Peer reviewed

Electricity Markets & Policy
Energy Analysis & Environmental Impacts Division
Lawrence Berkeley National Laboratory

Winners are not keepers: Characterizing household engagement, gains, and energy patterns in demand response using machine learning in the United States

Annika Todd-Blick, C. Anna Spurlock, Ling Jin, Peter Cappers, Sam Borgeson, Dan Fredman, Jarett Zuboy

December 2020

This is a pre-print version of an article published in *Energy Research & Social Science*.
DOI: https://doi.org/10.1016/j.erss.2020.101595

This work was supported by the U.S. Department of Energy Solar Energy Technologies Office under Lawrence Berkeley National Laboratory Contract No. DE-AC02-05CH11231.

Winners are not keepers: Characterizing household engagement, gains, and energy patterns in demand response using machine learning in the United States

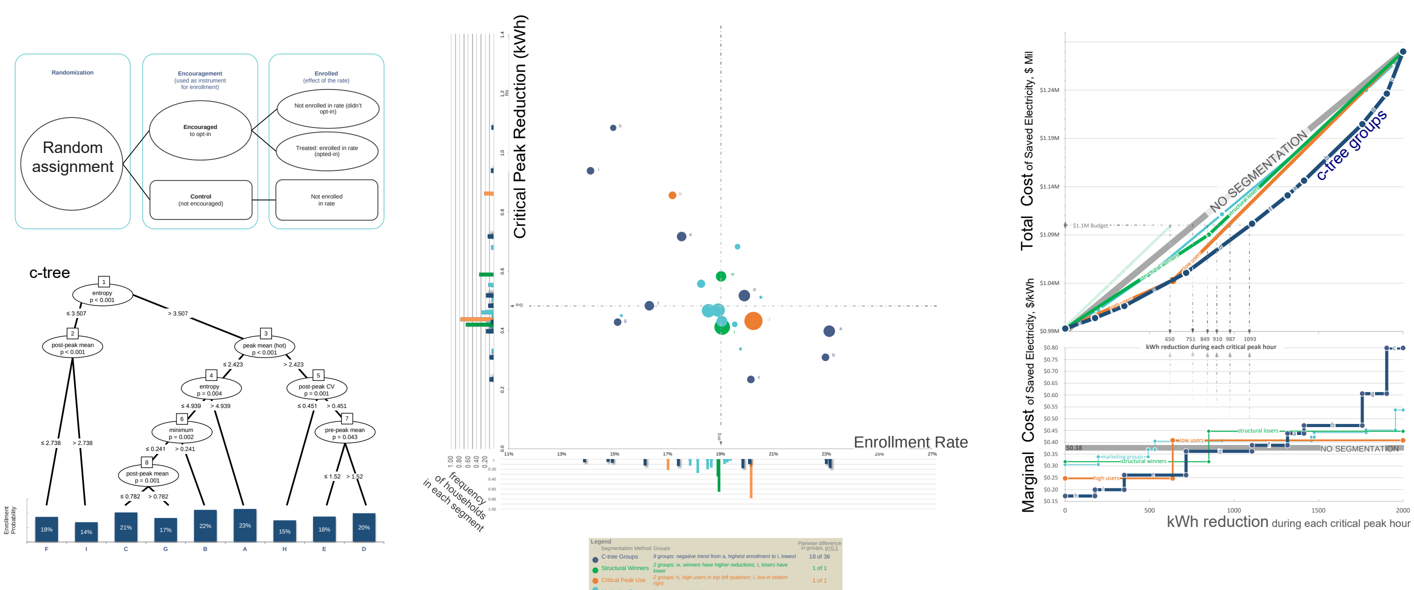
Annika Todd-Blick*, C. Anna Spurlock, Ling Jin, Peter Cappers, Sam Borgeson, Dan Fredman, Jarett Zuboy

*Lawrence Berkeley National Laboratory, Mail Stop: 90R4000, 1 Cyclotron Road, Berkeley, CA 94720, USA. e-mail address: ATodd@lbl.gov

A B S T R A C T

Demand-response programs can help utilities manage rapidly evolving electric grids, but these programs are subject to the complexities of human behavior. This paper explores a novel method for uncovering heterogeneity in households. We use a machine-learning method known as a Conditional Inference Tree (c-tree) algorithm to categorize households based on their energy behavior characteristics collected via smart meters, and explore how this translates through into heterogeneity in their real-world response to a DR program. Using data from randomized controlled trial, we generate estimates of the changes in energy use caused by the program within each household group. Our results show that the c-tree approach differentiates households by their energy-use characteristics in a way that increases the spread in enrollment rates and critical peak reduction among household groups, compared with the spreads achieved via several conventional segmentation methods. Thus, the c-tree analysis enables the most tailored targeting of major potential energy savers and could provide the greatest increase in cost-effectiveness of household recruitment into DR programs. Our results also offer fresh insights into the relationships between household energy behavior characteristics – such as peak energy use and "structural winningness" (the ability to save money under a DR program without changing energy-use beha-viors) – and household decisions about enrolling in DR programs and reducing energy use. Our research also demonstrates the potential of smart meter data, combined with machine learning and econometric methods, to provide significant value to utilities, program implementers, researchers, and other stakeholders.

G R A P H I C A L   A B S T R A C T



1. Introduction

Electric grids are experiencing a rapid increase in complexity due to growing demand, expansion of variable renewable energy resources (e.g., solar and wind), and the resulting increase in net electricity de-mand variability (e.g., increased deployment of distributed solar pho-tovoltaic systems, uncontrolled charging of electrified transportation)[1]. Supply-side investments have historically been used by utilities to meet such challenges, but they can be costly and take considerable time to implement [2]. By contrast, demand-side solutions – such as re-sidential demand response (DR), energy efficiency (EE), and behavior-based programs (BB) – can be implemented quickly at lower incre-mental expense, and they can be used on an as-needed basis [3]. For example, Consolidated Edison Company of New York (ConEd) elected to pursue a portfolio of demand-side solutions in 2014, including EE, DR, and distributed battery storage, with an authorized total budget of US$ 200 M to ensure the deferral of US$ 1.2B upgrade to an existing substation through 2026 [4].

However, demand-side programs rely on people making changes to their energy consumption behaviors. Because different types of con-sumers make energy-use decisions in different ways [5], household responses to these programs can vary widely. This variability and un-certainty can create challenges for program implementers and system operators/planners responsible for maintaining reliability; this can re-sult in a preference for supply-side resources [6,7].

Correlating household characteristics with energy-use decisions can reduce some of the uncertainty surrounding enrollment and/or per-formance of demand-side programs. In the hopes of gaining more consistent and dependable insights into which types of households are likely to consistently respond, electric utilities commonly group people based on sociodemographic data from surveys, census block informa-tion, past program participation, marketing categories (e.g., "young optimists," "success driven"), or other metrics that may correspond to a specific program's goals (e.g., segmenting high versus low peak energy users for a program that aims to reduce peak usage) [8–11]. They then may tailor program offerings or recruitment strategies to target groups that are deemed likely to provide more consistent and dependable energy or peak demand reductions. However, these methods can be costly and may include information about only a small subset of households (e.g., only respondents to survey instruments) and are not typically validated to determine whether the different segments of people actually respond differently [12–16].

New data emerging from the recent broad-based deployment of Advanced Meter Infrastructure ("smart meters") represent a direct, continuously updated source of information on hourly household energy use, improving significantly on the coarse monthly averages that were typically the only consumption data available in the past. The new smart-meter data are readily available as well as inexpensive to access and use (after the upfront cost of establishing algorithms and data routines). Using such readily accessible data to segment households by nuanced energy-consumption characteristics – such as high peak use, high-baseload/low-discretionary use, and use patterns – can help utilities more robustly and cost-effectively identify households that may be receptive and consistently responsive to demand-side programs.

Researchers are just beginning to exploit these more detailed energy-use data using innovative techniques such as machine-learning algorithms. Some have used daily electricity patterns to infer appliance ownership and predict a household's socio-demographic attributes [17–21]. Others have used patterns revealed from electricity load shapes to cluster households into daily use archetypes [22–25]. Still others summarize additional segmentation methods based on smart-meter data [18,26]. These studies posit that targeting certain household segments for utility demand-side programs is desirable because of higher energy-savings potential. However, few studies use real-world data to validate the hypothetical correlation between target groups and program uptake and/or energy savings.

In this paper, we use an analytical approach centered around real-world validation of such correlations. Instead of using conjecture to determine which groups might be receptive to demand-side programs, we use a machine-learning algorithm to categorize households based on their real-world responses to the programs and their pre-program energy-use characteristics collected via smart meters. We then estimate the changes in energy use caused by the demand-side program within each household group. To our knowledge, ours is the first study to verify how household groups created via machine learning correlate with real-world enrollment rates and energy-use changes in response to demand-side programs. Because we employ a unique dataset from a randomized controlled trial (RCT), the energy-use changes of each group due to the program are estimated with a high level of robust statistical confidence. We also compare the differentiation of household groups (in terms of enrollment and energy-use changes) achieved via our machine-learning approach – which is designed to maximize the differences between groups and the similarities within groups – versus the differentiation achieved via the type of marketing-categorization and simplistic energy-use approaches commonly used by utilities today. We then discuss how, on a practical level, the enhanced household-segmentation approach might be used to improve the cost-effectiveness of electricity DR programs. Finally, we enumerate a number of insights regarding underlying household behaviors with regard to their response to this pricing program that can inform how our results, specific to the setting of this particular utility and this particular demand-side program, might provide meaningful information that can be generalized beyond this setting.

Note that this is proof-of-concept research: we are illustrating the power of our combination of machine learning and smart meter data to generate evidence-based and data-driven insights about response to demand side programs that have not been uncovered in previous literature; compare segmentation using machine learning to other commonly used methods; and characterize households' energy behavior characteristics. Accordingly, the specific results reported may not necessarily be generalizable to other programs, other utilities, or other circumstances. The tools and methods we employ, however, can be replicated for these other situations.

2. Data and methods

2.1. Experimental design and data

We use a unique smart-meter dataset from pilot testing of a Sacramento Municipal Utility District (SMUD) time-based rate designed to shift electricity consumption away from peak-demand periods [27]. SMUD's time-of-use (TOU) rate charged higher prices[a] from 4 to 7 PM on all summer non-holiday weekdays and lower prices during other times. A critical peak pricing (CPP) rate charged higher prices[b] during peak hours on 23 critical event days over two years and lower prices during other times. The experimental prices were in effect between June 1 and September 30 in 2012 and 2013. Although we use some details from the CPP rate in our analysis, we focus on results for the TOU rate, because residential TOU programs are more prevalent than residential CPP programs among U.S. utilities today [28,29].

Our data consist of hourly energy use in kilowatt-hours (kWh) for almost 50,000 customers, including all households in our control group and two encouraged groups (one group encouraged to join a CPP rate and one group encouraged to join a TOU rate), regardless of whether they enrolled in the rate or opted out during the pilot period. These energy-use data were collected for one year before the start of the pilot period (June 1, 2011 – May 31, 2012) and for two summers during the pilot period (June 1, 2012 – September 30, 2012 and June 1, 2013 – September 30, 2013). This dataset is part of a much bigger dataset that includes additional pricing pilot arms, and a multitude of other variables, which we focused on in other papers[c].

Our methods draw from recent work on quantifying treatment-effect heterogeneity in RCTs [35,36]. The recruitment and enrollment methods for the SMUD pilots were implemented as a specific type of RCT called a randomized encouragement design (RED, Fig. 1). First,

[a] TOU rate option: participants were charged an on-peak price of $0.27/kWh between 4 and 7 PM on weekdays, excluding holidays. For all other hours, participants were charged $0.0846/kWh for the first 700 kWh in each billing period, with any additional use billed at $0.1660/kWh.

[b] Participants were charged $0.75/kWh during CPP event hours, when temperatures and SMUD's system loads were expected to be unusually high. This rate option was designed under the assumption that 12 CPP events would be called each year, between 4 and 7 PM on weekdays, excluding holidays. Customers were notified 24 h in advance of an event day. For all other hours, participants were charged $0.0851/kWh for the first 700 kWh in each billing period, with any additional use billed at $0.1665/kWh.

[c] Briefly, these papers: include a summary the data [7]; found that the overall effect of a TOU rate on peak energy reductions using a voluntary enrollment method was 11–13% but was on average 6% for those defaulted onto it [29,30]; found that critical peak pricing (CPP) rates resulted in electricity reductions on non-event days but during the same event hours [31]; found that critical peak event load reductions were similar between vulnerable (i.e., elderly, low income, and chronically ill) and non-vulnerable subpopulations [10,32]; found that relative to an opt-in enrollment method, a method that defaults customers onto a rate (but allows them to opt-out) results in much higher participation rates (98% versus 20%), roughly comparable attrition rates (3.9% vs. 4.4% for opt-out vs. opt-in respectively), and lower average energy reduction per enrollee (6% versus 17% for opt-out vs. opt-in respectively), which combines into higher overall savings if offered to the entire residential customer population (5.7% vs 3.3% aggregate peak period load reduction for opt-out vs opt-in respectively) [29,30]; and compares biases among various program evaluation methods relative to the estimates from the RCT, among other findings [33,34].

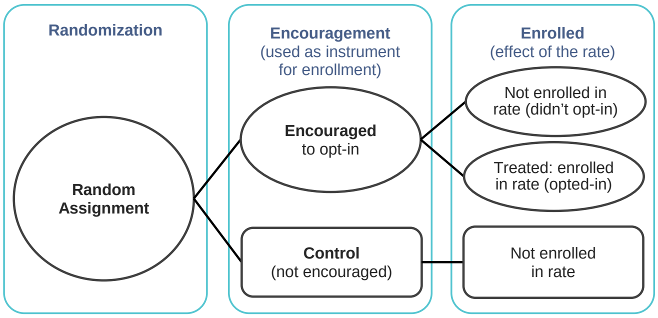


Fig. 1. RED and corresponding IV analysis method for default and voluntary enrollment rates.

households were randomly assigned to one of two groups[d]: the control group, which was not contacted and not encouraged to opt-in, and the encouraged group, which was contacted and encouraged to opt-in to the TOU rate through marketing materials. Households in the encouraged group could then choose to either do nothing, or to enroll in the rate. Those that chose to enroll in the rate are the "treated" group; i.e., their energy use was actually affected by the TOU rate. We provide multiple tests of validation of randomization in the Supplemental Files Appendix A; out of the many comparisons, the testing t-statistic exceeds one for only one, suggesting that the randomization yielded very similar groups.

Randomized implementations of these types of demand-side programs are rare, and they allow for particularly rich causal inference analysis and unbiased estimation of program results. This aspect of the data allows us, unlike other program evaluations, to create valid, credible, and unbiased estimates of program outcomes for various different segmenting approaches – thus enabling a valid, credible, and unbiased estimate of how the segmenting approaches compare. For example, some researchers have used daily electricity patterns to infer appliance ownership [17–20], to cluster households into daily use archetypes [22–25], or to create segmentation methods based on smart-meter data [26], but do not go further to evaluate how the inferred ownership or archetypes translates through to program response.

2.2. Heterogeneity: Household segmentation approaches

We compare several approaches to segmenting households into more homogeneous groups and assessing the heterogeneity between the groups with respect to enrollment and response. Our primary approach is based on a novel application of a machine-learning algorithm. We compare this with three additional approaches: one based on marketing data, another separating households by high versus low peak energy use, and a third separating households into structural winners versus losers. Structural winners are households that would receive lower bills on the rate if they did not change their energy use relative to the year prior to program implementation (i.e., while on the traditional time-invariant electricity rate). The first two are more likely to be used by utilities to examine heterogeneity within the population of customers. The third, although not commonly used by utilities to directly determine heterogeneity of enrollment and response between winners and losers, is often critically relevant to the discourse around implementing these types of rates. There is often times serious concern that allowing people to opt in to rates like these will result in large numbers of "free riders" (i.e., structural winners) joining the rate and simply saving money without providing peak period energy savings

2.2.1. Marketing categories

The marketing categories we use are those SMUD had defined for its customer base at the time of the pilot project [27]. We have no information about these categories other than the category names and the fact that SMUD used them for message targeting purposes in some cases. The groups are: Uninvolved Achievers; Green Boomers; Green Echoes; Money-Minded Strivers; Senior Savers; Young Families; Big Toys, Big Spender; Boomers and Buyers; and Unclassified.

2.2.2. High and low critical peak energy use

As another segmentation method, we simply order all households by critical peak energy use (defined as peak hour energy use on proxy critical peak days in the pretreatment period) and then split them into two groups at the 80th percentile value [37]. Thus, the high group represents the top 20% of critical peak energy users, and the low group represents the bottom 80%.

2.2.3. Structural winningness

Next, we segment households based on structural winningness. We calculate structural winningness in units of dollars per month during the summer, as the hypothetical difference between a household's energy bill in the summer before the program if it had been on the new rate and its energy bill in the same summer on the old rate—assuming the household's energy use remains the same under both rates. A positive structural winningness value means a household would hypothetically save money under the new rate without making energy-use changes, whereas a negative structural winningness value means a household would hypothetically lose money. Importantly, structural winningness does not represent the difference in actual bills from the period before to the period after the program.

For our analysis, we create two groups based on structural winningness. One group consists of all households that are structural winners, and the other consists of all households that are structural losers.

2.2.4. Machine learning (C-Tree Algorithm)

The innovation we contribute is the application of a machine learning approach, Conditional Inference Tree (c-tree) [38], to segment households into groups with similar energy characteristics and enrollment outcomes. The energy characteristics are derived from hourly electricity use data.

This approach, described in more detail in the Supplemental Files Appendix B, capitalizes on the insight that such trees can be effective at segmenting populations in econometric studies [39]. Note that the results provide the estimated likelihood of customer enrollment using only data observable prior to the start of the program. Unlike clustering algorithms, which are unsupervised machine learning methods, c-tree is supervised, and is more similar to a regression: its inputs include both explanatory variables (independent x variables), as well as an "explained" variable (a dependent y variable). Specifically, the c-tree algorithm identifies splits in explanatory variable values (i.e. segments customers into groups with values greater than or less than a specific value conditional on all preceding splits) that provide the most statistically significant difference in the explained variable across each group (see the Supplemental Materials for more details on the c-tree algorithm). The c-tree algorithm implements binary recursive partitioning based on the permutation test discussed in [40]. The stopping criteria in c-tree is based on hypothesis testing with a user defined significance level. Such a statistical approach aims to overcome the two fundamental problems found in other commonly used regression trees: overfitting and a selection bias towards covariates with many possible splits [38]. In contrast to other regression trees, where pruning procedures based on cross validation with a held-out subsample are often

[d] This description simplifies the full implementation. This pricing pilot was actually part of a much larger effort. The full details are described in [27]. Basically, out of the households that qualified for the study, they were randomly assigned (with a simple random sample) to one of ten groups that were separated. Nine of the groups were assigned to be encouraged to participate in nine variations of rates or combination of rates, while the tenth was held out and not contacted at all. There was no overlap between the groups in terms of participation.

Table 1
C-Tree Explanatory Variables: Household Behavioral Energy Characteristics and Their Proxy Metrics Related to Enrollment and Response.

Variability of a household's energy use schedule [13,26]	
Entropy	Day-to-day variability in overall energy-use patterns[i]
Pre-peak coefficient of variation (CV)	Day-to-day variability in use during the 2 h prior to the peak period
Peak CV	Day-to-day variability in use during peak period
Post-peak CV	Day-to-day variability in use during the 2 h following the peak period
Load magnitude during critical peak (hottest) days [10,14–16]	
Pre-peak mean (hot)	Average consumption during the 2 h prior to the peak period during the hottest days
Peak mean (hot)	Average consumption during the peak period during the hottest days
Post-peak mean (hot)	Average consumption during the 2 h following the peak period during the hottest days
Load magnitude during non-critical days [9] [k]	
Pre-peak mean	Average consumption during the 2 h prior to the peak period
Peak mean	Average consumption during the peak period
Post-peak mean	Average consumption during the 2 h following the peak period
Baseload use [3] [l]	
Minimum	Average daily minimum consumption across all days (i.e., base load)

[i] Generated by clustering daily discretionary use patterns and calculating the entropy in load shape assignment for a given customer across days. The load shape assignments and clusters were calculated in [33].
[k] The presence of residents during times surrounding the peak periods may make them more able to respond to time-varying rates.
[l] Baseload use may be less salient and therefore may not be fully accounted for when making energy decisions, but it does affect the bill-savings potential of each household.

used to avoid overfitting, the statistical procedures in c-tree ensure the right sized tree is grown without pruning or cross-validation.

In our case, the c-tree explanatory variables are various energy use metrics, and the explained variable is the choice of whether or not to enroll. To define the explanatory variables, we develop a set of behavioral energy characteristics that prior research and industry knowledge suggest might influence a household's willingness to enroll in and respond to time-varying pricing programs (defined in Table 1). We then calculate these characteristics for each household, using hourly smart-meter data from one year prior to the SMUD rate implementation (pre-treatment data).

While most of the explanatory variables in Table 1 are intuitive, the entropy variable deserves more explanation (further descriptions of all variables can be found in the Supplemental Files, Appendix C). The entropy variable is meant to characterize overall variability in the day-to-day consumption patterns of a given household. It is calculated as the degree to which a given household's daily load profile varies across all days. Specifically, we first applied adaptive k-means clustering to the daily load shapes of all customers and reduce them to a dictionary of 99 representative daily load shapes. For each household, for each day, we assign their load shape to one of those 99 representative shapes. The detailed procedure for this, including data preprocessing and hyper-parameter tuning, can be found in [33]. We then compute the daily load shape entropy for each household based on the occurrence frequency of their daily consumption patterns among the dictionary load shapes. For household n, the daily load shape entropy is computed as follows.

$$entropy_n = -\sum_{i=1}^{99} p_n(L_i)\ln(p_n(L_i))$$

where $p_n(L_i)$ is the occurrence frequency of the dictionary load shape L_i in household n. The high and low variation in the day-to-day consumption patterns of a given household can be then quantified according to the value of their load shape entropy. For example, for households with day to day consumption routines that do not vary, the above equation will yield an entropy value of zero because they can be represented by one dictionary load shapes with occurrence frequency of 100%. In contrast, households with daily usage patterns evenly distributed across all the 99 dictionary load shapes will have highest entropy value of 6.6.

The c-tree algorithm, applied to all households encouraged to enroll in either the CPP or TOU rates, ensures that the customer segmentation satisfies two objectives: (1) households within each segmentation group should have energy metrics that are as similar as possible to each other, and (2) the enrollment probabilities should span as wide a range as possible between the groups. We evaluate the heterogeneity of the enrollment probabilities of the resulting groups through pairwise hypothesis testing and the robustness of the resulting spread of the enrollment probability is confirmed by a random forest of 250c-trees (see Fig. 2 in the Supplemental Files, Appendix B).

2.3. Estimation of enrollment and reduction

2.3.1. Enrollment

As mentioned previously, we focus primarily on the TOU rate for the remainder of the analysis. For each of the groups created by any of the segmentation approaches, we calculate the percentage of the group that enrolled in the TOU program. This is a simple calculation of the number of households in each group that opted in to the new rate divided by the total number of households in the group.

2.3.2. Reduction in electricity use during critical peak hours

For each of the groups created by any of the segmentation approaches, we estimate the change in the most important periods of electricity use resulting from enrollment in the TOU rate. Specifically, we estimate the electricity "critical peak reductions" – the reductions in peak load hours on the hottest days, which are the most valuable and important to utilities in terms of grid reliability. We define these critical hours as they are defined under the CPP rate (see Section 2.1).

This estimate of the critical peak reduction can be accomplished because the characteristics that define the segments are also observed in the control group, so each group is made up of households that were in the control group, those that did not opt-into the rate when encouraged, and the treated group of households that opted into the rate when encouraged.

The appropriate method for estimation of the effect of the Treatment on the Treated (TOT) with a RED design is a difference-in-differences (DID) instrumental variable (IV) regression, which in this case is also called two-stage least squares (2SLS); see [41] for a detailed explanation. We use the randomized encouragement as the instrument for a household's enrollment in the rate (i.e., encouraged is an instrument for treated, which itself means the customer is placed on the rate). Through this regression, we estimate the critical peak reduction per critical peak hour, averaged over households within individual groups.

The estimation is shown in Eqs. (1) and (2), where y_{it} captures critical peak electricity consumption for household i in hour t; P_t is an indicator variable equal to one for all observations starting on June 1, 2012 (i.e., the treatment period) and thereafter, zero otherwise; T_{it} is an indicator variable equal to one for all observations starting on June 1, 2012 and thereafter if household i was enrolled in treatment on day t, zero otherwise; C_i^k represents a set of nine indicator variables, each equal to one if household i is in segment k, zero otherwise; φ and γ_i are household fixed effects; and ω_t and τ_t are hour-of-sample fixed effects. Data used are peak hours (4–7 PM) on the CPP days in each treatment summer as well as proxy-CPP days during the pretreatment summer (defined as weekdays with maximum hourly dry bulb temperature greater than or equal to the mean maximum hourly dry bulb temperature during CPP days during the treatment period, which was 96 degrees Fahrenheit). The first interaction term in the equation, $C_i^k \bullet P_t$ is a segment-specific treatment period indicator, equal to one for all observations starting on June 1, 2012 and thereafter if household i is in segment k, zero otherwise (note that this is needed because the household fixed effects plus the time fixed effects do not provide group specific time effects). We instrument for treatment T_{it} with randomized encouragement into the treatment indicator A_{it}, which is also equal to zero for all observations prior to the start of the treatment period (June 1, 2012), and equal to one for all hours t thereafter, so long as household i was randomized into the group that was encouraged into treatment. The coefficients of interest are the set of β_k parameters, which are energy savings as the difference between treated and control groups averaged over households within group k.

First stage:

$$T_{it} = \sum_{k=1}^{9} \eta_k (C_i^k \bullet P_t) + \sum_{k=1}^{9} \psi_k (C_i^k \bullet A_{it}) + \varphi_i + \omega_t + e_{it} \tag{1}$$

Second stage:

$$y_{it} = \sum_{k=1}^{9} \delta_k (C_i^k \bullet P_t) + \sum_{k=1}^{9} \beta_k (C_i^k \bullet \hat{T}_{it}) + \gamma_i + \tau_t + \varepsilon_{it} \tag{2}$$

Average critical peak reductions of all households combined are estimated with the DID IV specification in Eqs. (3) and (4), where T_{it} is instrumented by $A_{it,}$ and defined as above, and β is the coefficient of interest, which serves as a reference measure of average critical peak reduction in the absence of group heterogeneity.

First stage:

$$T_{it} = \eta \bullet A_{it} + \varphi_i + \omega_t + e_{i,} \tag{3}$$

Second stage:

$$y_{it} = \beta \bullet \hat{T}_{it} + \gamma_i + \tau_t + \varepsilon_i \tag{4}$$

2.4. Calculation of marginal and total costs per kWh of peak reduction

Incorporating the enrollment probability and the estimates of the critical peak reductions, together with the fixed costs of implementing the program and the cost of recruiting households, it is straightforward to calculate the total and marginal cost per unit of energy savings for each segmentation method, and thus their cost-effectiveness if used for targeting a subset of households for enrollment in the program.[e] These costs are calculated per unit of critical peak reduction for our four segmentation methods; one unit is the *average* reduction during *each and every* critical peak hour, in kilowatt hours (kWh) (i.e., to calculate the total reduction over the entire summer, one would need to multiply this quantity by the total number of critical peak hours over that summer). The marginal cost is the slope of the total cost curve, and represents the cost of obtaining one extra unit of reduction. We normalize based on the number of customers such that the total amount of reduction, if every customer were recruited, would be 2000kWh. For the total costs, we use program startup costs of $748,000, and program administration cost of $245,000, both taken from the utility-reported costs [27].

[e] The costs associated with developing the various groups for each segmentation method are not included in this analysis. Instead, we assumed the costs associated with such analyses were already sunk and thus not included as part of the marginal and total cost analysis.

3. Results

3.1. Results: Groups formed by the C-Tree segmentation approach

The results from the c-tree algorithm are displayed in Fig. 2, which provides a visualization of the recursively-partitioned tree. It shows which of the explanatory variables, along with the corresponding cutoff points, are chosen by the algorithm to best segment households. In this case, the metrics that are chosen by the algorithm include entropy, post-peak CV, peak mean (hot), minimum, pre-peak mean, and post-peak mean (as described in Table 1).

Fig. 2 illustrates how the segments are defined: starting with splitting the households based on their entropy level, and then hierarchically using each of the chosen variables and its cutoff to further split the households, resulting in the final c-tree segmentation groups A through I. Each household is assigned to one of these groups based on whether the household's energy characteristics shuttle it to the right or left branch of each node. For example, starting at node 1, if a household has an entropy metric of 3.507 or lower, that household continues on the left branch to node 2. From node 2, if that same household's post-peak mean use is greater than 2.738, then the household splits off to the right branch and becomes part of group E.

Fig. 3 shows the average characteristics of the nine household groups defined by the c-tree based off of the combined populations of those encouraged into either the TOU or CPP treatment. The groups range in enrollment rates from 14% (group I) to 23% (group A). The color gradient in Fig. 3 illustrates the differences in average metric values among the groups, ranging from dark grey (lowest values) to light grey (highest values). For example, group I has the lowest average entropy value (2.7, dark grey), and group A has the highest (5.2, light grey). The color pattern in Fig. 3 – or lack thereof – suggests both the complex relationships among household characteristics and enrollment rates as well as the c-tree's ability to account for this complexity when segmenting groups in a way that maximizes the differences among group enrollment rates. No metric exhibits a consistent pattern in values across columns (i.e., when scanning rows from left to right, low enrollment to high enrollment).

3.2. Results: Heterogeneity of enrollment rate and critical peak reduction across household segments for each approach

We now derive the enrollment rate and estimate the critical peak reduction for each group within each of the four segmentation methods for the TOU encouraged group only. Table 2 shows the results of our analysis. The middle columns show the enrollment rate for each group defined by each of the four segmentation methods (c-tree, marketing, high vs. low critical peak period use levels, and structural winners vs. losers) when the sample is limited to the group encouraged to enroll in TOU only. The rightmost columns show the estimated treatment-on-the-treated effect, in terms of the critical peak reductions (under the TOU rate) for each of the groups within each method. To examine, for each method, how the groups created by that method differ from each other, we present Table 3, which reports the p-values from a pairwise comparison of whether the estimate for each group is statistically

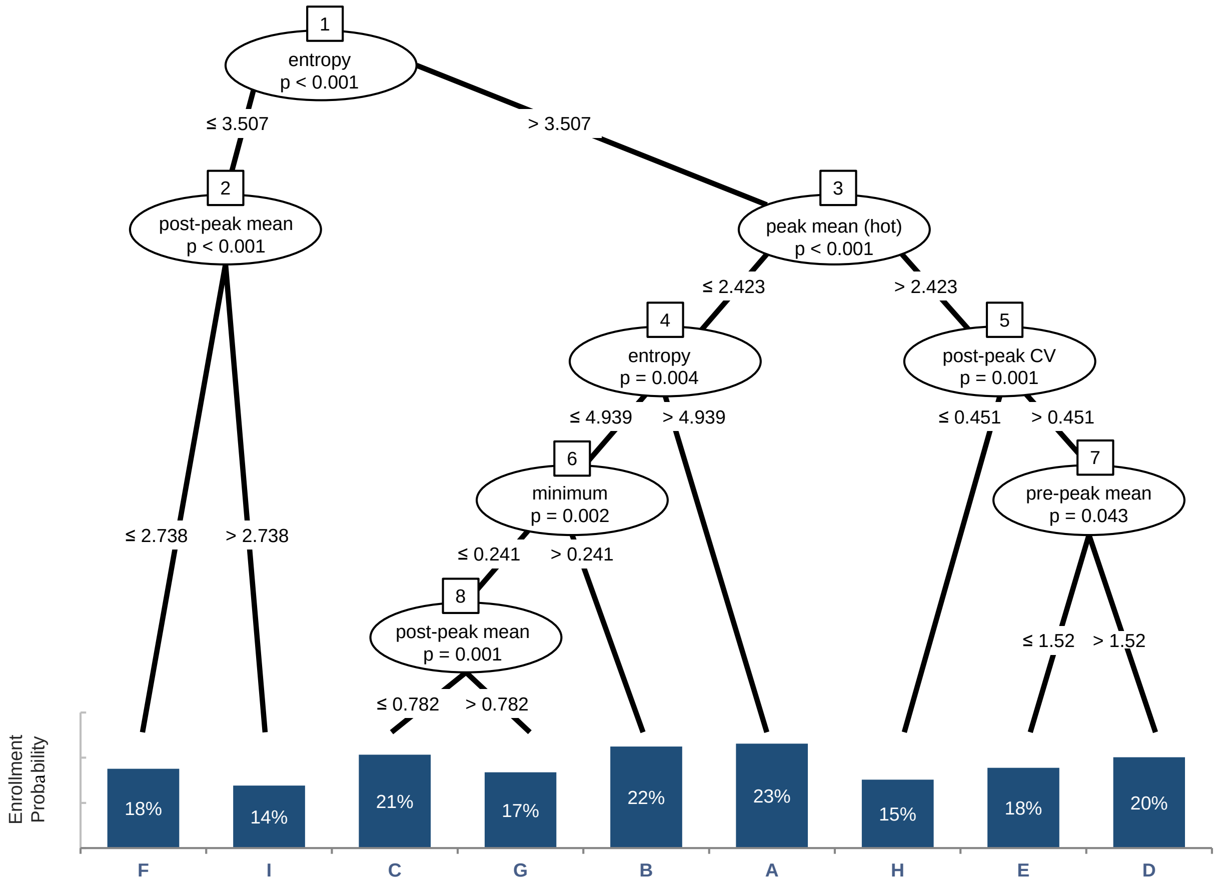


Fig. 2. Results of the c-tree algorithm defining household groups A–I by cutoffs in household energy behavior metrics for the TOU and CPP encouraged groups. *The chosen variables are labeled in the nodes; their corresponding cutoff values are shown in the branches below each node; the p-value for the probability that the split formed by that cutoff for that variable results in a wider spread of groups is shown in the node; the end-point node (*i.e.*, leaves of the tree) shows the label we gave to each group (A through I), and the enrollment probability for each of the groups.*

	I (14%)	H (15%)	G (17%)	F (18%)	E (18%)	D (20%)	C (21%)	B (22%)	A (23%)	Range
Variability of a household's energy use schedule										
Entropy	2.7	4.0	4.4	2.8	4.7	4.4	4.4	4.4	5.2	2.7 to 5.2
Post-peak CV	0.4	0.4	0.7	0.5	0.7	0.6	0.9	0.6	0.8	0.4 to 0.9
Load magnitude during critical peak (hottest) days										
Peak mean (hot)	5.3	3.9	1.6	2.5	3.2	4.0	0.8	1.6	1.4	0.8 to 5.3
Load magnitude during non-critical days										
Pre-peak mean	3.5	2.1	0.8	1.6	1.1	2.2	0.4	1.0	0.7	0.4 to 3.5
Post-peak mean	3.8	2.8	1.2	1.5	1.8	2.4	0.5	1.3	1.0	0.5 to 3.8
Baseload Use										
Minimum	0.8	0.6	0.2	0.3	0.3	0.5	0.1	0.4	0.2	0.1 to 0.8

C-tree Customer Segments (Enrollment %)

Fig. 3. Average metrics for each c-tree-derived household group, including the six behavioral energy characteristics that define the groups.

significant from the other groups' estimates (statistically significant differences are shown as colored cells).

For the c-tree groups, the range of critical peak reduction estimates is 0.23–1.08 kWh/h per household, and the enrollment percentages range from 14% to 23%. As shown in Table 3, many of the differences in critical peak reduction between the c-tree groups are statistically significant; out of 36 total pairwise differences, exactly half (18) are statistically significant ($p < 0.10$).

For the nine marketing groups, the range of critical peak reduction estimates is much narrower: 0.33–0.68 kWh/h. The enrollment percentages range is also narrower: 15%–20%. None of the pairwise differences in critical peak reduction between marketing groups are statistically significant, suggesting that using marketing groups provides no real differentiation of households that is meaningful for energy savings resulting from the program.

For the two groups segmented into high and low monthly critical

Table 2

Critical Peak Reductions for Each of the Four Segmentation Techniques For Households Enrolled in the TOU Group. *Notes: The last two columns show the treatment-on-the-treated (TOT) estimator (in this case, the treated are the households enrolled, and the standard errors (se).*

				Effect of treatment on the enrolled, average reduction per household per critical peak hour (kWh)	
Ctree groups	**N encouraged**	**N (treated)**	**% of encouraged that enrolled**	**TOT**	***se***
A	3073	710	23.1%	−0.40**	*0.06*
B	1668	383	23.0%	−0.31**	*0.08*
C	1678	338	20.1%	−0.23**	*0.07*
D	3101	617	19.9%	−0.52**	*0.10*
E	2379	417	17.5%	−0.72**	*0.11*
F	2183	356	16.3%	−0.48**	*0.11*
G	1231	186	15.1%	−0.43**	*0.12*
H	890	133	14.9%	−1.08**	*0.19*
I	1072	151	14.1%	−0.94**	*0.26*
Total	17275	3291	19.1%	−0.47**	*0.04*
Marketing groups	**N encouraged**	**N enrolled (treated)**	**% of encouraged that enrolled**	**TOT**	***se***
Big Toys, Big Spender	535	104	19.4%	−0.33	*0.20*
Boomers, Buyers Green	187	28	15.0%	−0.44	*0.43*
Boomers Green Echoes	297	60	20.2%	−0.51*	*0.24*
Money Minded Strivers	2260	406	18.0%	−0.55**	*0.10*
Senior Savers	3494	650	18.6%	−0.46**	*0.08*
Uninvolved Achievers	1654	318	19.2%	−0.41**	*0.11*
Young Families	1133	219	19.3%	−0.68**	*0.14*
Unclassified	2924	548	18.7%	−0.42**	*0.09*
Total	4791	958	18.2%	−0.46**	*0.07*
	17275	3291	19.1%	−0.47**	*0.04*
Monthly critical peak usage level groups	**N encouraged**	**N enrolled (treated)**	**% of encouraged that enrolled**	**TOT**	***se***
High	2297	394	17.2%	−0.86**	*0.00*
Low	8232	1663	20.2%	−0.44**	*0.00*
Total	10529	2057	19.5%	−0.48**	*0.04*
Structural winningness groups	**N encouraged**	**N enrolled (treated)**	**% of encouraged that enrolled**	**TOT**	***se***
Structural winners	5827	1106	19.0%	−0.59**	*0.00*
Structural losers	11045	2101	19.0%	−0.42**	*0.00*
Total	16872	3207	19.0%	−0.47**	*0.04*

peak use, the critical peak reduction estimates are 0.44 kWh/h for low and 0.86 kWh/h for high, and the enrollment percentages are 17% for low users and 20% for high users, ranges that are both narrower than those observed across the c-tree groups. The difference in critical peak reduction between the two groups is statistically significant ($p = 0.002$).

Finally, for the two groups segmented by structural winningness, the critical peak reduction estimates are 0.42 kWh/h for structural losers and 0.59 kWh/h for structural winners (the narrowest of all ranges analyzed), and the enrollment percentages are almost identical at 19%. The difference in critical peak reduction between the two groups is statistically significant ($p = 0.001$).

3.3. Results: Differences in outcomes between groups within each segmentation methods for the TOU rate

Fig. 4(a) visually shows the spread of the groups in terms of critical peak reduction and enrollment percent for each group within the four segmentation methods – c-tree, marketing, high vs. low monthly peak period use, and structural winners vs. losers. The histograms bordering the axes show the frequency of households assigned to each group through the c-tree algorithm (i.e., from all of the households that were encouraged, regardless of whether they enrolled or not). Fig. 4(b) shows the enrollment and critical peak reduction for each group; this is a visual representation of the estimates shown in Table 2. The graph shows the correlation between critical peak reduction (on the y axis, in kWh per hour), and enrollment rate. It shows this correlation as a bubble on the graph for each group within each of the four segmentation methods that we are examining, where the size of the bubbles equate to the number of households in the group. Specifically, the nine groups in the marketing segmentation approach are shown with turquoise dots; the two groups, low and high critical peak use, are shown with orange dots; the two groups of structural winners versus structural losers are shown in green; and the nine groups derived from the c-tree algorithm are shown in blue.

This graph can be interpreted as a visual display of the spread, across enrollment rates and critical peak reductions, among each of the groups for each of the four methods. A method that has a wider spread indicates that the groups formed by that method differ more across the relevant dimension. At a high level, it is apparent that the blue c-tree groups are more widely spread than the other methods' groups. Fig. 4(b) summarizes the enrollment and critical peak reduction ranges for each segmentation method, the number of pairwise tests of each group relative to other groups within a segmentation method that are statistically significant, and the Chi-squared statistic.

More specifically, the results show that of the four segmentation methods, the c-tree approach produces the greatest differentiation among groups in terms of both average enrollment rates (9 percentage point difference between groups with the minimum enrollment rate of 14% and maximum of 23%) and critical peak reduction estimates (0.86 kWh/h difference between groups with minimum reduction of 0.23 kWh/h and maximum of 1.08 kWh/h). The nine marketing groups produce a smaller range of both enrollment rates (5 percentage points) and critical peak reduction estimates (0.34 kWh/h); recall as well that none of the energy savings differences between these marketing-based groups are statistically significant. Compared with the marketing

Table 3
P-values from Pairwise Tests of Statistically Significant Differences between Each Group Relative to Every Other Group within one Segmentation Method, and the F statistic for the Chi squared test of explanatory power of heterogeneity, for Critical Peak Load Reduction Estimates. Performed for: (a) the 9 groups in the C-tree Segmentation Method, (b) the high vs. low peak use Segments; (c) the Marketing Groups; and (d) the Structural Winners vs. Losers.

(a) Ctree groups: 18 of 36 pairs are statistically significantly different at p > 0.1

	F	I	C	G	B	A	H	E	D
F		0.105	0.049	0.731	0.203	0.482	0.007	0.123	0.814
I	0.105		0.008	0.074	0.021	0.042	0.654	0.429	0.129
C	0.049	0.008		0.155	0.482	0.059	0.000	0.000	0.016
G	0.731	0.074	0.155		0.415	0.818	0.004	0.070	0.562
B	0.203	0.021	0.482	0.415		0.382	0.000	0.003	0.106
A	0.482	0.042	0.059	0.818	0.382		0.001	0.008	0.289
H	0.007	0.654	0.000	0.004	0.000	0.001		0.096	0.009
E	0.123	0.429	0.000	0.070	0.003	0.008	0.096		0.167
D	0.814	0.129	0.016	0.562	0.106	0.289	0.009	0.167	

F(17, 50905) = 218.08 Prob > F = 0.0000

(b) Marketing groups: 0 of 36 pairs are statistically significantly different at p > 0.1

	Big Toys, Big Spender	Boomers, Buyers	Green Boomers	Green Echoes	Money Minded Strivers	Senior Savers	Uninvolved Achievers
Big Toys, Big Spender		0.810	0.565	0.315	0.535	0.709	0.148
Boomers, Buyers	0.810		0.899	0.806	0.966	0.946	0.604
Green Boomers	0.565	0.899		0.858	0.860	0.722	0.529
Green Echoes	0.315	0.806	0.858		0.497	0.357	0.468
Money Minded Strivers	0.535	0.966	0.860	0.497		0.725	0.185
Senior Savers	0.709	0.946	0.722	0.357	0.725		0.137
Uninvolved Achievers	0.148	0.604	0.529	0.468	0.185	0.137	
Young Families	0.660	0.964	0.743	0.339	0.751	0.941	0.122
Unclassified	0.531	0.970	0.852	0.462	0.986	0.721	0.165

F(17, 50905) = 10.66 Prob > F = 0.00

(c) High vs. Low use: statistically significantly different at p > 0.1

	High	Low
High		0.002
Low	0.002	

F(3, 50905) = 521.38
Prob > F = 0.0000

(d) Structural Winners vs. Losers: statistically significantly different at p > 0.1

	Winners	Losers
Winners		0.001
Losers	0.001	

F(3, 50905) = 68.40
Prob > F = 0.0000

Note: P-values shown, cells are colored when differences are statistically significant ($p < 0.10$).

groups, the two critical peak use segments generate an even smaller range in enrollment rates (3 percentage points); their range in critical peak reduction estimates (0.42 kWh/h) is larger, but it is still smaller than the c-tree group range. Finally, segmenting customers based on structural bill impacts (structural winners versus structural losers) produces the smallest ranges of enrollment rates (almost identical) and critical peak reduction estimates (0.16 kWh/h).

3.4. Results: Marginal and total costs per kWh of peak reduction

In this section, we simulate the application of various segmentation methods to recruitment into a TOU rate to compare each method's cost-effectiveness when used for targeting purposes (i.e., recruiting only certain groups of households in cases with program or portfolio budget constraints, and/or tailoring marketing messages for specific types of households).

Fig. 5 illustrates the total program cost (top) and marginal cost (bottom) curves for our four segmentation methods. The unit of the x-axis is the average critical peak reduction (kWh/h). The marginal cost is the slope of the total cost, and represents the cost of obtaining one extra unit of reduction.

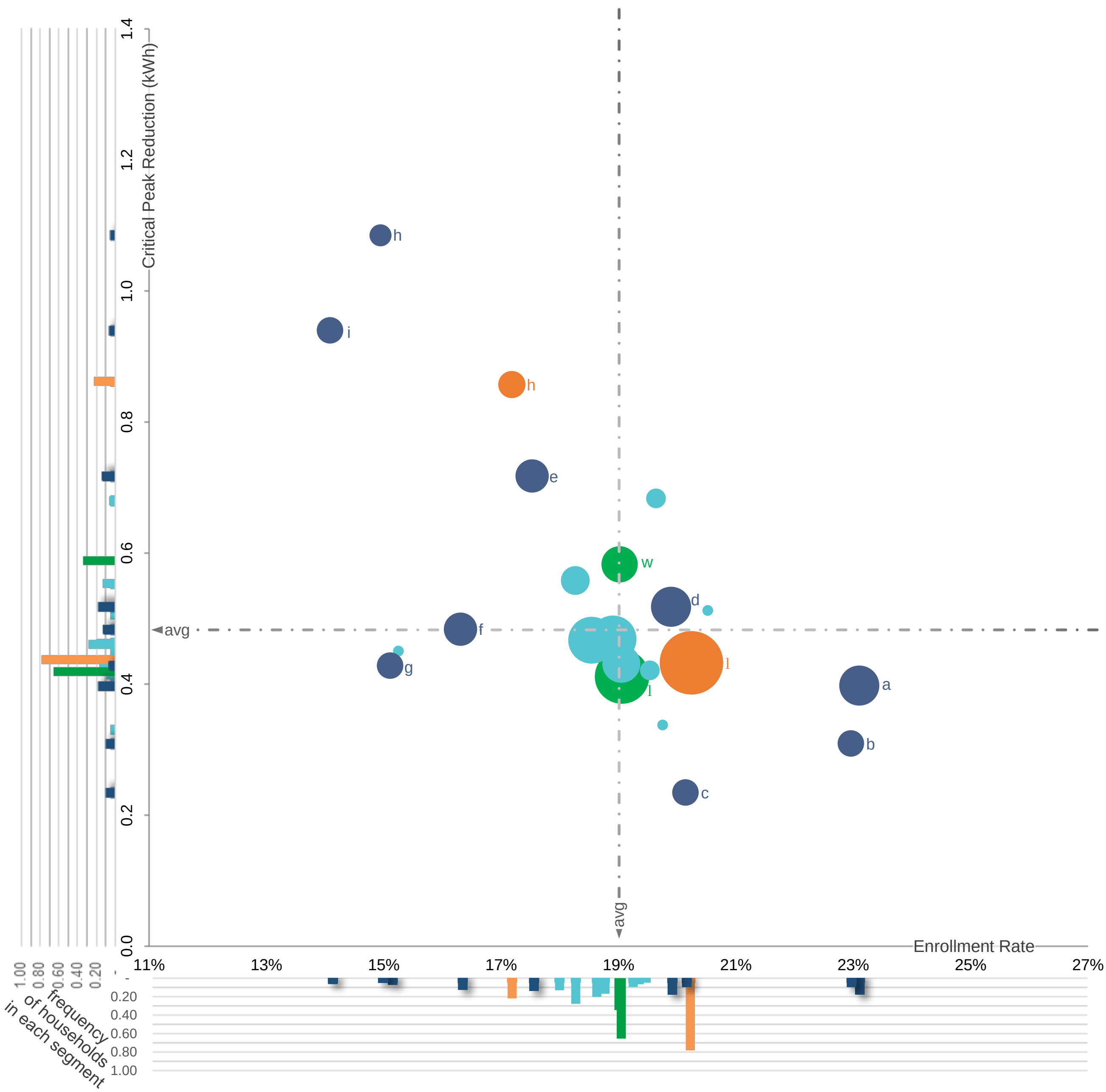
These figures should be interpreted in the following way. For the bottom graph on marginal cost, the grey dashed line is flat, which reflects a reference where each extra kWh/h critical peak reduction costs exactly the same amount, regardless of how much load response has already occurred. In contrast, the curves with a positive slope indicate that there is a method for ordering household segments such that the cheapest critical peak reductions can be obtained first, followed by the second cheapest, and so forth, ending with the most expensive critical peak reductions. A steeper slope indicates that the segmentation method depicted can detect households in such a way that the first set of critical peak reductions are significantly cheaper, capturing more savings for programs with limited budgets. The top graph of total cost contains the same information, but in a way that shows what total critical peak reductions can be obtained for each specific budget amount. A total cost curve that is more concave indicates that at *each* incrementally higher level of budget, more critical peak reductions can be obtained from that segmentation method.

Using the segmenting methods for ordering marketing and enrollment efforts is not helpful if the goal is to obtain all available critical peak reductions. In that case, all four segmentation methods cost the same amount to enroll all customers (i.e., they all have the same total budget of \$1.28 M to procure 2000 kWh/h of critical peak reduction). However, if instead the utility has a limited budget or only wants to enroll customers up to the point where the cost to acquire is no greater than the benefits that inure from their participation (i.e., a cost effectiveness target)[f], then Fig. 4(a) can be used to show how the segmentation methods differ in how much critical peak reductions each provides based on budget constraint (i.e., upper graph) or at the margin (i.e., lower graph).

Specifically, examining the marginal cost, with no segmentation (i.e., if every customer in this population were recruited), the average

[f] Although rarely done, utilities can set their assigned payment rate under CPP equal to the utility's marginal avoided cost. In those instances, the application of segmentation methods for recruitment would ensure that a utility only recruited customers up to the point where the cost to acquire the next one was less than the value they provided.

(a)



(b)

Legend

Segmentation Method	Groups	Pairwise difference in groups, p<0.1	Enrollment Rate min, max	Enrollment Rate Range	Peak Reduction min, max	Peak Reduction Range
C-tree Groups	*9 groups: negative trend from* a, *highest enrollment to* i, *lowest*	18 of 36	*14%, 23%*	**9%**	*0.23, 1.08*	**0.86**
Structural Winners	*2 groups:* w, *winners have higher reductions;* l, *losers have lower*	1 of 1	*19%, 19%*	**0%**	*0.42, 0.59*	**0.16**
Critical Peak Use	*2 groups:* h, *high users in top left quadrant;* l, *low in bottom right*	1 of 1	*17%, 20%*	**3%**	*0.44, 0.86*	**0.42**
Marketing Groups	*9 groups (see table 3b)*	0 of 36	*15%, 20%*	**5%**	*0.33, 0.68*	**0.34**

Fig. 4. (a) Critical peak reduction vs. enrollment rate for groups defined by each segmentation approach. (b) Ranges of Enrollment Rates and Critical Peak Reductions for Segmentation methods.

cost per unit of critical peak reduction as a whole is $0.38/kWh (grey dotted line). The c-tree approach (blue line) provides the greatest resolution and range in marginal cost per unit of critical peak reduction: the most cost effective group H is much lower, at only $0.17/kWh. The two least cost effective groups, C and G, cost $0.79/kWh and $0.61/kWh, respectively; eliminating (i.e., not recruiting) these two groups would significantly decrease the average cost of kWh reduction by 24%, from $0.38/kWh down to $0.29/kWh. In contrast, for the high/low

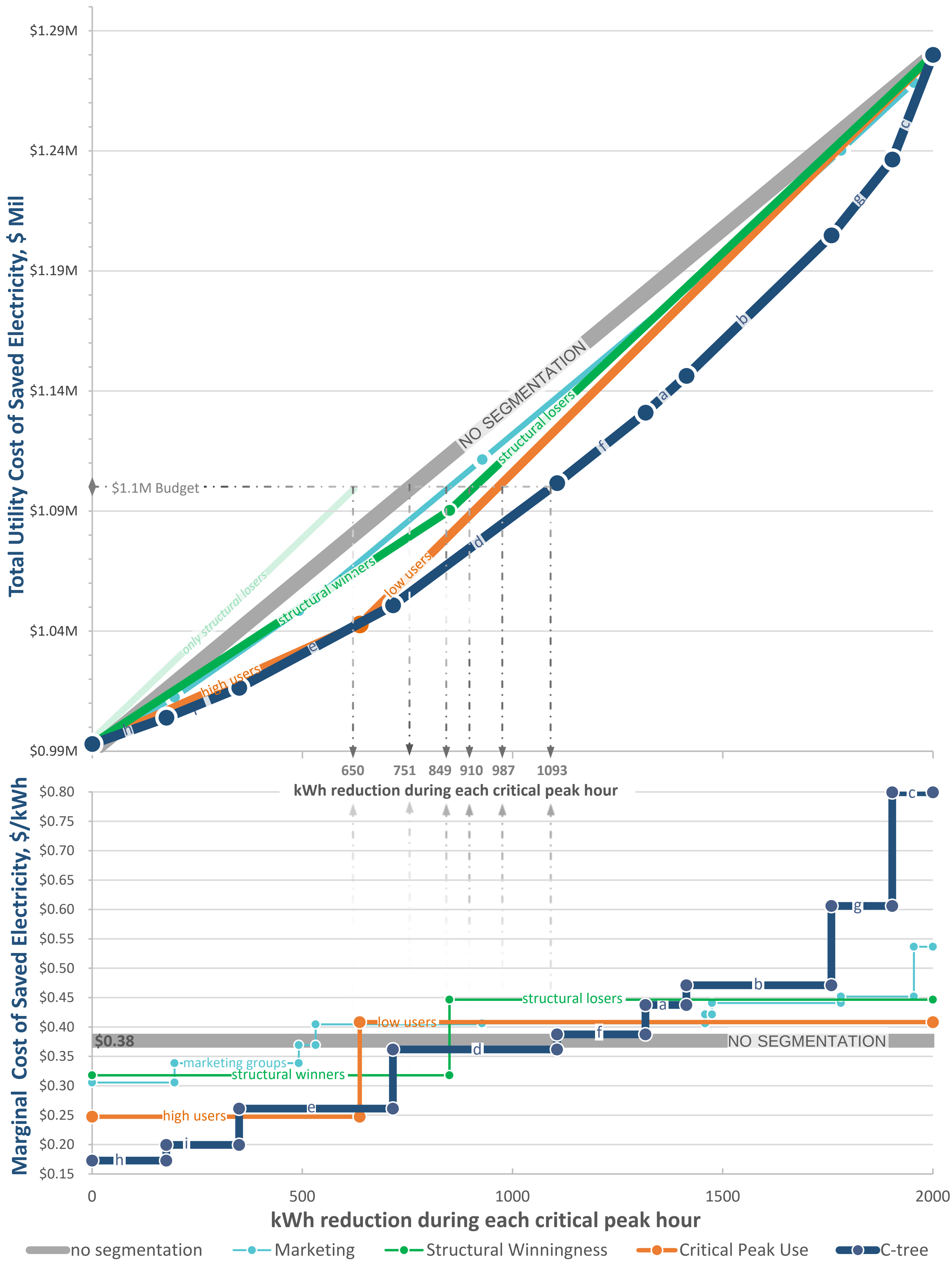


Fig. 5. Total cost (top) and marginal cost (bottom) per average critical peak kWh reduction.

usage segmentation approach (orange line), the reductions from the most cost-effective group, the high users, costs \$0.22/kWh – more ex-pensive than the c-tree group, but cheaper than groups from other segmentation approaches. The marketing groups have a lower spread in marginal cost, ranging from \$0.28/kWh to \$0.57/kWh. The structural winningness groups (green line) provide the least differentiation in marginal cost, from \$0.32/kWh to \$0.45/kWh.

Examining the total cost, with no segmentation (i.e., if every customer in this population were recruited), the total reduction would be 2000 kWh/h, and the total cost would be \$1.28 Million. Now consider the case of a utility with a fixed program budget of \$1.1 million. Without segmentation, the utility would scale back the number of customers in such a way that there would only be 751 kWh/h of critical peak reduction (see the top panel of Fig. 5). With segmentation, this utility could target and recruit the most cost-effective households first, resulting in a much higher level of critical peak reduction. Using the c-tree segmentation method to target households with the lowest marginal cost, this would result in a 46% increase in critical peak reduction, to 1093 kWh/h (for recruiting groups H, I, E, and most of D). In contrast, a utility that segmented customers by marketing groups would only result in 849 kWh/h of critical peak reductions for the same total utility budget, only a 13% improvement over no segmentation at all. By target marketing based purely on average critical peak use, the utility would first attract all of the high users and enroll roughly a quarter of the low users, who would collectively provide 987 kWh/h of critical peak re-ductions. This simple approach would produce 34% more reductions in critical peak load. Segmentation based on structural winningness is only more cost-effective than no segmentation if the correct group is recruited first. In our study, targeting structural *winners* would be more cost-effective, and would be pursued first along with some of the structural losers, producing 910 kWh/h of critical peak reductions for the same total utility budget. However, if utilities avoided households they knew were structural winners during recruitment out of concern that they would not provide any critical peak reductions once on the rate, the result would be *worse* than not segmenting at all – the utility's program would only be able to generate ~650 kWh/h of critical peak reductions.

4. Discussion

Efforts to understand and reduce household heterogeneity – how different groups of customers enroll and respond differently to a demand-side program – may be useful for utilities to pursue in several ways. Here we discuss a few insights that are gained from examining the heterogeneity produced by each of the four segmentation methods.

4.1. Tradeoffs and usefulness of segmentation methods

4.1.1. The biggest differentiation between groups versus the easiest method

The results show that of the four segmentation methods, the c-tree approach produces the greatest differentiation amongst groups in terms of both average enrollment rates and critical peak reduction estimates. This means that this method will produce more nuance for understanding household energy decisions, and more options in terms of targeting for enrollment in different demand-side programs. However, the c-tree method also requires an analysis using a statistical package with machine learning algorithms, requires the ability of an analyst to understand and perform such algorithms, and requires storing hourly data for every household. While these analytic abilities are very easy to implement once they are appropriately coded into a utility's automatic software, the upfront cost of training current staff, or hiring or outsourcing experts in data science may be prohibitive. There are some open source codes and algorithms that are specifically designed to make this procedure easier and in line with current industry standards, although they are not plug-and-play and require data skills and time to get up and running).[g]

In contrast, segmenting households into high and low critical peak hours performs quite well in terms of differentiating critical peak reductions, although it does not do a good job at separating the enrollment rate. Data and analysis required to place households into one of these two groups is straightforward. It only requires recording data for each household for a few peak hours during the handful of critical event days, rather than every hour on every day. Furthermore, these hours can be averaged across critical peak events for each customer, so that only one metric per household needs to be actually stored. Follow-on research outside the scope of this paper could go deeper into segmentation methods that use only data from the critical peak hours, to refine and optimize the cutoffs for groups. For example, perhaps good differentiation could result from separating households into more groups and/or using a different cutoff. In addition, research could explore a continuum of tradeoffs between the level of differentiation versus analytics (perhaps in terms of the storage and skills needed to perform the algorithms), so that utilities and other program administrators could weigh the benefits and costs of implementation depending on their particular circumstances.

4.1.2. Improving cost effectiveness through segmentation

In a world with constrained budgets, where utilities and other program administrators need to optimize allocation of resources to the most cost-effective programs, and/or meet cost-effectiveness requirements, an important value of segmenting is the ability to target the program offering specifically to groups of households that provide the biggest return – that is in the context of reliability demand-side programs, those that provide the cheapest critical peak electricity or demand reductions. If utilities focused their recruitment efforts on households that are likely to have the lowest cost per unit of energy savings – i.e., targeted only the households in the best performing groups – they could increase the cost-effectiveness of demand-side programs, particularly when households can be directed to the best fit among multiple non-overlapping programs.

Because the c-tree analysis provides the highest resolution on household group characteristics, it enables the most tailored targeting of major potential energy savers and thus could provide the greatest increase in cost-effectiveness of recruitment. However, there is one rather large additional drawback to the c-tree method as applied in this proof-of-concept example: the c-tree method requires a program-relevant outcome variable (such as the choice of whether or not to enroll, as in our case). Results from our c-tree are not directly generalizable to other utilities. We are not using the c-tree method to *predict* program enrollment, but simply as a statistically-driven method to define heterogeneous groups of households based on multiple energy use features. Therefore, other utilities and other programs wishing to apply a similar method would need to conduct their own analysis. This could be done if a program were first piloted, and results from that pilot were analyzed using a method similar to the c-tree method we define, the results of which could then be used to guide recruitment priorities for a full-scale program rollout. However, one of the additional contributions of our analysis is that segmenting households using a data-driven method such as a c-tree provides an opportunity to generate insights regarding the nature of the underlying behaviors, which could inform application of targeting, using a variety of methods, to other settings. These additional insights are discussed in the next section.

4.2. Insights into energy behaviors

In addition to showing how the segmentation methods perform

[g] For example, VISDOM is an open source code specifically made for utilities to to use to pull out useful energy characteristics out of hourly smart meter data [42].

based on their ability to generate groups with meaningful differentiation, the results provide information about how people make energy decisions. This information may challenge perceptions commonly held by utilities about which types of customers are willing to enroll in, and are able to respond to, demand response programs.[h]

Utilities and other program implementers have historically used very coarse, if any, form of segmentation for targeting and optimizing enrollment in demand-side programs (e.g., high energy users, zip code, yearly bill; see [43,44] for example). This is despite the fact that segmentation is so prolifically used in many other areas; the Obama campaign was famously successful in using improved voter choice prediction methodologies to inform micro-targeting techniques for optimizing voter solicitation [45]. In addition, online companies such as Google, Facebook, and Amazon, among others, have built empires on effectively micro-targeting advertisements and product suggestions across their user/customer base. If utilities with limited program budgets could target only certain groups of households that are more likely to provide the sought-after system benefit or grid service, they could get a bigger response per program dollar spent. Another use that could help increase cost effectiveness would be to provide specialized, tailored marketing messages. The observations described in the remainder of this section can provide some insights that could inform potentially useful approaches for some of this tailoring or recruitment targeting.

4.2.1. Enrollment and critical peak reduction have a negative correlation

Based on our experience, there is a common assumption that the households more likely to respond to a demand-side program (by generating critical peak reductions) are also more likely to enroll in the program. The intuition behind this assumption is that people make their enrollment decision based on the extent to which they perceive they could adapt to the rate. If they perceive they have a low cost of reducing their own consumption in the peak periods, they may be more likely to enroll. However, the enrollment rate appears to correlate negatively with critical peak reductions (Fig. 4(a)). In other words, households that are *less* likely to enroll end up providing *larger* reductions once they are enrolled, and vice versa. This trend runs contrary to the common assumption.

4.2.2. Households with high critical peak use provide more energy reduction

Utilities frequently contend that households with high critical peak use can provide a larger critical peak load response to the demand-side program, simply because they have more load to give. Our analysis supports this assertion, namely, that high-peak-use customers reduce their critical peak use more than low-use customers (see the orange dots in Fig. 4(a); this difference is statistically significant).

4.2.3. Structural winners are not free riders: They provide larger critical peak reductions

If a household is told, or is able to derive for themselves, the fact that they're a structural winner who would save money on the new rate without making any changes to their energy behavior[i], there is a common concern among program implementers and economists, who theorize that households know their own use patterns and optimize accordingly [46,47], that they will game the system by enrolling in demand-side programs and will then provide disproportionately less, or even no, electricity reduction than their structural "loser" counterpart once enrolled (i.e., act as free riders). However, our results suggest that in this case structural winners are *not* free riders – they actually provide substantially *larger* (~40% more) critical peak reductions once enrolled (equivalent to 19 percentage points in Fig. 4(a); this difference is statistically significant). It is not clear why this is the case; there is the chance that these households do not correctly calculate the financial savings they would experience from the program, or they do and realize they could gain even further savings by changing their behavior.

4.2.4. The usefulness of marketing categories in this context is uncertain

In this study, the marketing categories were developed by the utility for the broader purpose of understanding their customers, and not to optimally segment customers for these specific rate pilots. And while we would therefore expect that any method specifically tailored for the TOU rate would perform better, we still hypothesized that the marketing categories would be somewhat useful in differentiating groups with higher or lower enrollment and/or response. In contrast to these expectations, the marketing segmentation approach seems to shed no light on which types of customers respond: the differences between marketing groups are small, and none of the differences are statistically significant. This suggests that such an approach may offer limited insights into exploring heterogeneity in enrollment probability or responsiveness to new electricity rates. However, these marketing groups may be meaningful for many other purposes within the rate offering, such as tailoring messages to encourage program enrollment, or outside of the rate offering. Our study does not consider these purposes; all customers encouraged to enroll in the demand-side program at SMUD received the same recruitment messages.

4.2.5. Energy-use behaviors are complex

The three more traditional customer-segmentation approaches that we consider are based on prior beliefs about energy behaviors, which our findings may confirm (high users *do* produce greater peak reduction), contradict (structural winners enroll *at the same rate* and provide *greater* peak reduction), or render inconsequential (marketing-group segmentation does not meaningfully differentiate customer enrollment or peak reduction). In contrast, the machine learning c-tree segmentation approach is based on algorithms that sift through data to find the mix of variables that create the largest spread across groups. While the set of variables that are inputs into the c-tree algorithm are based on prior beliefs about which characteristics may impact enrollment and response, there are no prior beliefs and no structure imposed on which of those variables the algorithm deems are important for splitting the households into segments, and in which direction the correlation (if any) is. Yet the c-tree method performs better than the other methods in that it differentiates households by their energy-use characteristics in a way that increases the spread in enrollment rates and critical peak reduction among household groups compared to other segmentation methods, demonstrating the presence of complex, multifaceted energy-use behaviors.

The patterns resulting from the c-tree algorithm output offer insights into household energy behaviors. In particular, the c-tree group definitions align somewhat with the other segmentation methods. For example, groups H and I are characterized by high structural winningness and high critical peak energy use, and – as results from the other segmentation methods would suggest – these groups provide the largest per-household critical peak reductions once enrolled (Fig. 4(a)).

It is also interesting to examine the primary differentiating factor in the c-tree groups: entropy, or variation in energy use patterns across days. At this point we can only speculate as to why entropy is a primary differentiating factor; perhaps households with similar patterns of energy behavior every day (low entropy) have a varying routine of work

[h] The authors have been supporting utilities, regulators, and policymakers in the implementation, execution, and evaluation of demand-side programs collectively for over 30 person-years. Preconceived notions and perceptions of customers' willingness to enroll and ability to respond are rarely formally included in written documents. Instead, they are more often revealed through conversations. Thus, in this section when we refer to "common assumptions," they are based on our own experiences with utilities, policymakers, regulators, and electric industry stakeholders over the course of our collective careers.

[i] Many households classified as structural winners have both high peak and high baseload use. They are structural winners because their monetary loss from higher on-peak prices is overwhelmed by the monetary gain from the slightly lower off-peak prices charged by this program against their larger amount of off-peak baseload.

and home time but have a programmed thermostat, or they may have a set work and home routine that results in the same climate-control actions every day (such as turning on the air conditioner every day at 5:00 PM when they get home). Or perhaps entropy is simply a proxy, which correlates well with an entirely different energy characteristic that we do not capture.

This illustrates the power of smart-meter data combined with machine learning: it enables identification of nuanced but meaningful distinctions of energy behaviors among household groups, which open up avenues for deeper exploration of the behavioral mechanisms underlying these patterns.

5. Conclusions

The data from SMUD's time-based rate pilot program, which was designed as a randomized control trial, enables a rigorous analysis of the connections among household characteristics, program enrollment, and resulting energy-use changes. By understanding the pre-program household characteristics and program results, we can retroactively examine – with an unusually high level of statistical confidence – the validity of segmenting households based on a range from traditional to more novel machine-learning approaches.

Our results show that the c-tree machine-learning approach differentiates households by their energy-use characteristics in a way that increases the spread in enrollment rates and critical peak reduction among household groups compared to other segmentation methods. Specifically, the spread in enrollment and critical peak reductions for the c-tree-defined groups are, respectively, 9 percentage points and 0.9 kWh/h; this is substantially higher than the spreads for the marketing-based groups (5 percentage points and 0.3 kWh/h), a simple approach based on high and low critical peak use groups (3 percentage points and 0.4 kWh/h), and an approach based on the structural winningness of households (0 percentage points and 0.2 kWh/h). Statistical analysis suggests that the c-tree, high/low, and structural winningness segmentation methods can produce groups with significantly different critical peak reductions, whereas the marketing-based approach cannot.

Our results also reveal interesting insights into the nature of how residential electricity consumers make decisions, and how these decisions compare to conventional wisdom. In line with common beliefs, households that use more energy during critical peak hours produce more energy reductions. Households that are more likely to enroll do *not* produce the most reductions. Directly in opposition to conventional wisdom, households that are structural winners do *not* enroll more than others, and when they do enroll, they are not free-riders; in fact, they produce higher critical peak reductions than structural losers do. Finally, the c-tree segmentation illustrates the power of smart-meter data combined with machine learning: it enables identification of nuanced but meaningful distinctions of energy behaviors among household groups that would not necessarily emerge from intuition and assumptions.

Because the c-tree analysis provides the highest resolution on household group characteristics, it enables the most tailored targeting of major potential energy savers and thus could provide the greatest increase in cost-effectiveness of recruitment. For example, a utility constrained to a $1.1 million annual budget could enroll customers using the c-tree segmentation approach and produce 46% more aggregate critical peak reductions than it could using an approach that does not segment customers at all. However, segmenting customers by high and low critical peak use produces reduction that are 34% more than no segmentation, and this approach requires no machine-learning capabilities, minimal data, and does not necessitate observation of a program outcome variable, such as choice of whether or not to enroll. The upfront cost of cleaning the data and implementing the machine-learning analysis should be considered before choosing a segmentation method. It is unclear whether marketing segmentation would provide any benefit, because there are no statistically significant increases in peak reduction from no segmentation method. Segmenting by structural winningness could be beneficial, as in our case it would produce a 21% greater reduction than no segmentation, but only if the non-intuitive group (the structural winners) were targeted first. This segmentation should therefore be approached with caution depending on the setting.

These findings are specific to the demand-side programs implemented by SMUD, and they cannot be directly applied to programs instituted by other utilities. However, our analysis demonstrates, in a general way, the value of segmenting customers for enrollment efforts in demand-side programs, and more specifically illustrates the potential value of combining smart-meter data with machine learning to clarify the household characteristics relevant to energy-use behaviors and inform decision making. Similar methods could be applied to other types of energy programs as well, such as those aimed at deploying distributed solar energy systems, advanced transportation, or energy-efficiency measures, as well as being used to improve predictions of program outcomes.

Funding

This work was supported by the U.S. Department of Energy Solar Energy Technologies Office under Lawrence Berkeley National Laboratory Contract No. DE-AC02-05CH11231. This manuscript has been authored by an author at Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231 with the U.S. Department of Energy. The U.S. Government retains, and the publisher, by accepting the article for publication, acknowledges, that the U.S. Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for U.S. Government purposes.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] E. Marris, Upgrading the grid: Electricity grids must cope with rising demand and complexity in a changing world. Emma Marris explores the intricacies involved in controlling the power supply, Nature 454 (7204) (2008) 570–574.
[2] J. Wilkerson, P. Larsen, G. Barbose, Survey of Western US electric utility resource plans, Energy Policy 66 (2014) 90–103.
[3] A. Satchwell, R. Hledik, Analytical frameworks to incorporate demand response in long-term resource planning, Utilities Policy 28 (2014) 73–81.
[4] ConEd, BQDM Quarterly Expenditures & Program Report, Consolidated Edison Company of New York. Filed with the New York Department of Public Service in Case Number 14-E-0302., 2019.
[5] B. Parrish, P. Heptonstall, R. Gross, B.K. Sovacool, A systematic review of motivations, enablers and barriers for consumer engagement with residential demand response, Energy Policy 138 (2020) 111221.
[6] P. Cappers, C. Goldman, D. Kathan, Demand response in US electricity markets: Empirical evidence, Energy 35 (4) (2010) 1526–1535.
[7] P. Cappers, R. Scheer, American Recovery and Reinvestment Act of 2009: Final Report on Customer Acceptance Retention, and Response to Time-Based Rates from Consumer Behavior Studies, in: Lawrence Berkeley National Laboratory (Ed.) Berkeley, CA, 2016.
[8] B.A. Smith, J. Wong, R. Rajagopal, A simple way to use interval data to segment residential customers for energy efficiency and demand response program targeting, ACEEE Summer Study on Energy Efficiency in Buildings, sn, 2012, pp. 374-386.[9]
P. Alstone, J. Potter, M.A. Piette, P. Schwartz, M.A. Berger, L.N. Dunn, S.J. Smith, M.D. Sohn, A. Aghajanzadeh, S. Stensson, 2025 California Demand Response Potential Study-Charting California's Demand Response Future. Final Report on

Phase 2 Results, Lawrence Berkeley National Laboratory, Berkeley, CA, 2017.
[10] P. Cappers, C.A. Spurlock, A. Todd, J. Ling, Are vulnerable customers any different than their peers when exposed to critical peak pricing: Evidence from the U.S, Energy Policy 123 (2018) (2018) 421–432.
[11] M. Hayn, V. Bertsch, W. Fichtner, Electricity load profiles in Europe: The importance of household segmentation, Energy Res. Social Sci. 3 (2014) 30–45.
[12] S.J. Moss, M. Cubed, K. Fleisher, Market segmentation and energy efficiency program design, California Institute for Energy and Environment, Berkeley, 2008.[13] L. Dethman, D. Thomley, Comparison of segmentation plans for residential energy consumers, Energy Trust of Oregon Inc, 2009.
[14] P. Durand, Market,, segmentation: Defining a new attitude for utilities, Smart Grid Consumer Collaborative (2011).
[15] Smart Grid, Consumer Collaborative, Excellence in consumer engagement, Presentation (2011).
[16] D. Frankel, S. Heck, H. Tai, Using a consumer-segmentation approach to make energy-efficiency gains in the residential market, McKinsey and Company, 2013.
[17] A. Albert, R. Rajagopal, Smart meter driven segmentation: What your consumption says about you, IEEE Trans. Power Systems 28 (4) (2013) 4019–4030.
[18] A. Albert, R. Rajagopal, Thermal profiling of residential energy use, IEEE Trans. Power Systems 30 (2) (2014) 602–611.
[19] F. McLoughlin, A. Duffy, M. Conlon, A clustering approach to domestic electricity load profile characterisation using smart metering data, Appl. Energy 141 (2015) 190–199.
[20] J.P. Gouveia, J. Seixas, Unraveling electricity consumption profiles in households through clusters: Combining smart meters and door-to-door surveys, Energy Build. 116 (2016) 666–676.
[21] R. Razavi, A. Gharipour, Rethinking the privacy of the smart grid: What your smart meter data can reveal about your household in Ireland, Energy Res. Social Sci. 44 (2018) 312–323.
[22] I. Dent, U. Aickelin, T. Rodden, T. Craig, Finding the creatures of habit; Clustering households based on their flexibility in using electricity 1 (2012) 2012.
[23] H.-Â. Cao, C. Beckel, T. Staake, Are domestic load profiles stable over time? An attempt to identify target households for demand side management campaigns, IECON 2013–39th Annual Conference of the IEEE Industrial Electronics Society, IEEE (2013) 4733–4738.
[24] J. Kwac, R. Rajagopal, Demand response targeting using big data, IEEE International Conference on Big Data, IEEE, 2013, pp. 683–690.
[25] J. Kwac, J. Flora, R. Rajagopal, Household energy consumption segmentation using hourly data, IEEE Trans. Smart Grid 5 (1) (2014) 420–430.
[26] Y. Wang, Q. Chen, C. Kang, M. Zhang, K. Wang, Y. Zhao, Load profiling and its application to demand response: A review, Tsinghua Sci. Technol. 20 (2) (2015) 117–129.
[27] J.M. Potter, S.S. George, L.R. Jimenez, SmartPricing Options Final Evaluation, in: Sacramento Municipal Utility District (Ed.) U.S. Department of Energy, 2014.
[28] R. Hledik, A. Faruqui, C. Warner, The National Landscape of Residential TOU Rates, in: The Brattle Group (Ed.) 2017.
[29] M. Fowlie, C. Wolfram, C.A. Spurlock, A. Todd, P. Baylis, P. Cappers, Default Effects and Follow-on Behavior: Evidence from an Electricity Pricing Program, Under resubmission at RESTUD (2017).
[30] P. Cappers, C.A. Spurlock, A. Todd, P. Baylis, M. Fowlie, C. Wolfram, Time-of-Use as a Default Rate for Residential Customers: Issues and Insights, Lawrence Berkeley National Laboratory (Ed.), Berkeley, CA, 2016.
[31] A. Todd, P. Cappers, A. Spurlock, J. Ling, Spillover as a cause of bias in baseline evaluation methods for demand response programs, Appl. Energy 250 (2019) (2019) 344–357.
[32] P. Cappers, C.A. Spurlock, A. Todd, J. Ling, Experiences of Vulnerable Residential Customer Subpopulations with Critical Peak Pricing, Lawrence Berkeley National Laboratory (Ed.), Berkeley, CA, 2016.
[33] L. Jin, C.A. Spurlock, S. Borgeson, D. Fredman, L. Hans, S. Patel, A. Todd, Load Shape Clustering Using Residential Smart Meter Data: A Technical Memorandum, Lawrence Berkeley National Laboratory, 2016.
[34] L. Jin, D. Lee, A. Sim, S. Borgeson, C.A. Spurlock, D. Fredman, L. Hans, S. Patel, A. Todd, Comparison of Clustering Techniques for Residential Energy Behavior using Smart Meter Data, AAAI Workshops - Artificial Intelligence for Smart Grids and Buildings, San Francisco, CA, 2017.
[35] K. Imai, M. Ratkovic, Estimating treatment effect heterogeneity in randomized program evaluation, Ann. Appl. Stat. 7 (1) (2013) 443–470.
[36] S. Athey, G.W. Imbens, Machine learning methods for estimating heterogeneous causal effects, Stat 1050 (5) (2015).
[37] A. Todd, M. Perry, B. Smith, M. Sullivan, P. Cappers, C. Goldman, Insights from Smart Meters: Identifying Specific Actions, Behaviors and Characteristics that Drive Savings in Behavior-Based Programs, Lawrence Berkeley National Laboratory (Ed.), 2014, p. 14.
[38] T. Hothorn, K. Hornik, A. Zeileis, Unbiased recursive partitioning: A conditional inference framework, J. Comput. Graphical Stat. 15 (3) (2006) 651–674.
[39] T.K. Ho, Random decision forests, Proceedings of 3rd international conference on document analysis and recognition, IEEE, 1995, pp. 278–282.
[40] H. Strasser, C. Weber, On the asymptotic theory of permutation, Statistics (1999).
[41] P. Cappers, A. Todd, M. Perry, B. Neenan, R. Boisvert, Quantifying the Impacts of Time-based Rates, Enabling Technology, and Other Treatments in Consumer Behavior Studies: Protocols and Guidelines, Lawrence Berkeley National Laboratory (Ed.), Berkeley, CA, 2013, p. 115.
[42] J. Flora, T. Chin-Woo, S. Borgeson, R. Rajagopal, Stanford Energy Visualization and Insight System for Demand Operations and Management (VISDOM), 2017. https://tomkat.stanford.edu/innovation-transfer/visdom. (Accessed July 2018.
[43] BGE, Delmarva Power, PEPCO, Time-of-Use (TOU) Pilot Customer Education Plan: Maryland, Public Conference 44 (PC44): Transforming Maryland's Electric Grid, Maryland Public Service Commission (2018).
[44] Energy Central, How segmentation helps utilities engage small commercial customers, 2014. https://www.energycentral.com/c/iu/how-segmentation-helps-utilities-engage-small-commercial-customers. (Accessed November 27, 2019.
[45] G.R. Murray, A. Scime, Microtargeting and electorate segmentation: data mining the American National Election Studies, J. Political Market. 9 (3) (2010) 143–166.
[46] L.A. Skumatz, M.S. Khawaja, J. Colby, Lessons learned and next steps in energy efficiency measurement and attribution: energy savings, net to gross, non-energy benefits, and persistence of energy efficiency behavior, Prepared by Skumatz Economic, Energy and Environment, Research Associates for California Institute for, Berkeley, CA, 2009.
[47] R. Hartway, S. Price, C.-K. Woo, Smart meter, customer choice and profitable time-of-use rate option, Energy 24 (10) (1999) 895–903.

Supplemental Files for:

Winners are not keepers: Characterizing household engagement, gains, and energy patterns in demand response using machine learning in the United States

Appendix A. Broader Context and Background of Pricing Study, Design of Trial Implementation and Randomization Methods

The trial described in this paper was part of a much larger trial with multiple arms, which itself was part of a series of studies. Details follow.

Broader Context

Time-based rates[a] provide an opportunity for customers and utilities alike to achieve a variety of benefits including increased opportunity for customer bill management, lower utility power production costs, deferred future generation investments, and increased utilization of existing infrastructure. Historically, implementation of time-based rates required replacement of a traditional bulk usage electro-mechanical meter with either a multi-register electro-mechanical meter or an electronic interval meter that was accompanied by a monthly meter charge. The costs of individual meter upgrades was seen by many as a barrier to broader adoption of time-based rates. Recent broad-based deployment of Advanced Metering Infrastructure (AMI) removes this metering hurdle, thereby enabling the opportunity for broader adoption of time-based rates. Currently, utilities in the U.S. have installed more than 50 million smart meters, covering over 43% of U.S. homes (Institute for Electric Innovation, 2014).

The American Recovery and Reinvestment Act of 2009 included $3.4B for the Smart Grid Investment Grant (SGIG) program with the goal of creating jobs and accelerating the transformation of the nation's electric system by promoting investments in smarter grid technologies, tools and techniques (DOE, 2012). Among other topics, the Funding Opportunity Announcement (DE-FOA-0000058) identified interest in AMI projects that examined the impacts and benefits of time-based rate programs and enabling control and information technologies through the use of randomized controlled experimental designs.

[a] Time-based rates capture temporal differences in the cost of providing electricity. Some time-based rate designs are static, where the price schedule of electricity is set months, if not years, ahead of time to capture the diurnal and/or seasonal differences in cost (e.g., time-of-use pricing). Other time-based rate designs are more dynamic, where the price schedule is set 24 hours or less ahead of time based on anticipated or actual power system conditions, wholesale power costs, or both (e.g., critical peak pricing, variable peak pricing, real-time pricing).

Consumer Behavior Studies: Mandate for Randomized Controlled Experiments

Based on responses to this FOA, DOE decided to co-fund ten utilities to undertake eleven experimentally-designed Consumer Behavior Studies (CBS) that proposed to examine a wide range of the topics of interest to the electric utility industry. Each chosen utility was to design, implement and evaluate their own study in order to address questions of interest both to itself and to its applicable regulatory authority, whose approval was generally necessary for the study to proceed. The DOE Office of Energy Delivery and Electricity Reliability (OE), however, did set guidelines, both in the FOA and subsequently during the contracting period, for what would constitute an acceptable study under the Grant.

To assist in ensuring these guidelines were adhered to, OE requested that LBNL act as project manager for these Consumer Behavior Studies to achieve consistency of experimental design and adherence to data collection and reporting protocols across the ten utilities. As part of its role, LBNL formed technical advisory groups (TAG) to separately assist each of the utilities by providing technical assistance in all aspects of the design, implementation and evaluation of their studies. LBNL was also given a unique opportunity to perform a comprehensive, cross-study analysis that uses the customer-level interval meter and demographic data made available by these utilities due to SGIG-imposed reporting requirements, in order to analyze critical policy issues associated with AMI-enabled rates and control/information technology. Over the next several years, LBNL will publish the results of these analyses in a series of research reports that attempt to address critical policy issues relating to on a variety of topics including customer acceptance, retention and load response to time-based rates and various forms of enabling control and information technologies.

SMUD's Consumer Behavior Study

Through the U.S. Department of Energy's (DOE) Smart Grid Investment Grant Program (SGIG), the Sacramento Municipal Utility District (SMUD) designed and implemented a Consumer Behavior Study (CBS) of voluntary and default TOU rates that provide useful information and insights for addressing some of the key unresolved issues concerning a transition to default residential TOU rates.[b]

[b] See Appendix A for more background on the SGIG consumer behavior study effort and Appendix B for more details about SMUD's consumer behavior study.

SMUD conducted one of the largest and most extensive consumer behavior studies under the SGIG program. One of the study's main goals was to better understand how the enrollment approach (voluntary vs. default) affected enrollment rates, drop-out rates, and electricity demand impacts associated with time-based rates. SMUD implemented three different time-based rate designs, all in effect during the summer months (June to September) of 2012 and 2013: (1) a two-period TOU rate with a three-hour (4-7 p.m.) peak period, (2) CPP overlaid on an underlying tiered rate, and (3) CPP overlaid on the TOU rate (see Figure A1 and Table A1).[c] Like most of the other consumer behavior studies implemented under the SGIG program, SMUD's study utilized a true experimental design (i.e., randomized control trial and randomized encouragement design) in order to more credibly and precisely estimate the load response to these various rates. In this paper, only the customers included in the default TOU rate with IHD offer and opt-in TOU rate with IHD offer cells will be analyzed and discussed.

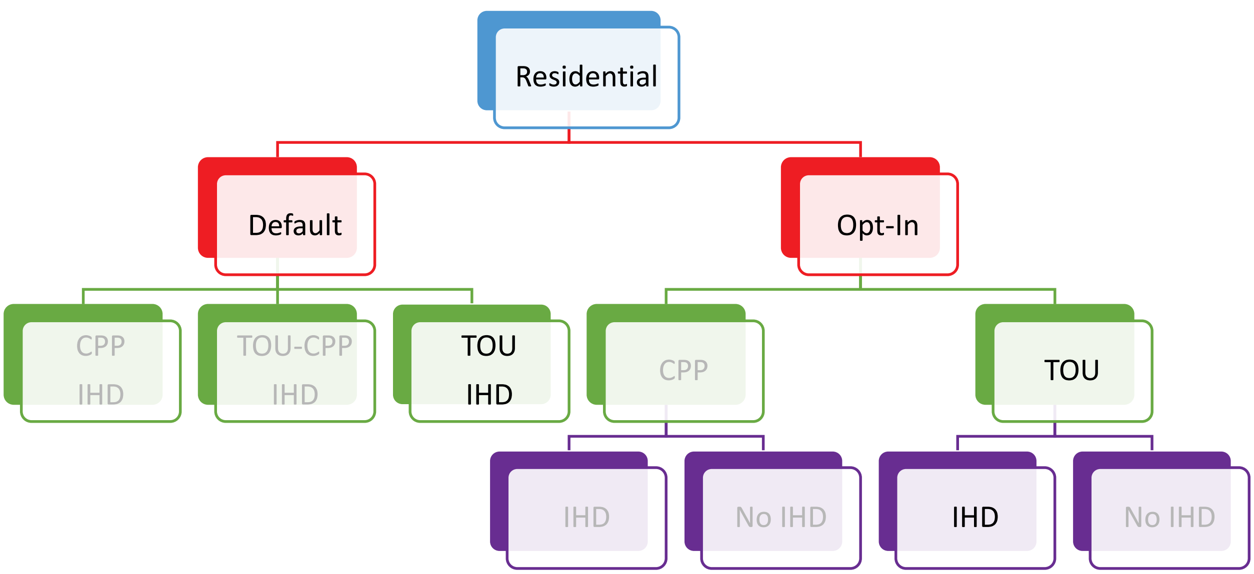


Figure A1. SMUD's Consumer Behavior Study Experimental Design

[c] Only the TOU and CPP were implemented in such a way that the effect of enrollment approach (voluntary vs. default) could be analyzed.

Period	CPP	TOU	TOU-CPP
Base (< 700 kWh)	8.51		
Base (> 700 kWh)	16.65		
Off-Peak (< 700 kWh)		8.46	7.21
Off-Peak (>700 kWh)		16.60	14.11
Peak		27.00	27.00
Critical Peak	75.00		75.00

Table A1. SMUD’s CBS Rate Design (¢/kWh)[d]

[d] Study participant on SMUD’s Energy Assistance Program (EAPR) rate faced different prices than those listed in Table 1.

Design of Trial Implementation and Randomization Methods

Overview

Sacramento Municipal Utility District (SMUD) is a summer peaking municipal electric utility with ~625,000 customers in its ~900 square mile service territory that covers much of the Sacramento, CA metropolitan area. SMUD's SGIG project (SmartSacramento) includes a consumer behavior study that evaluates customer acceptance and response to enabling technology combined with various time-based rates under different recruitment methods. The utility is targeting AMI-enabled residential customers across the entire service territory to participate in the study.

Goals and Objectives

This study focuses on evaluating the timing and magnitude of changes in residential customers' peak demand patterns due to exposure to varying combinations of enabling technology, different recruitment methods (i.e., opt-in vs. opt-out), and several time-based rates. SMUD is also interested in learning about customer acceptance of the different time-based rates under the alternative recruitment methods.

Treatments of Interest

Rate treatments include the implementation of three time-based rate programs in effect from June through September: a two-period TOU rate that includes a three-hour on-peak period (4 - 7 p.m.) each non-holiday weekday; a CPP overlaid on their underlying tiered rate; and a TOU with CPP overlay (TOU w/CPP) (see Table A1). Customers participating in any CPP rate treatments receive day-ahead notice of critical peak events, called when wholesale market prices are expected to be very high and/or when system emergency conditions are anticipated to arise. CPP participants will be exposed to 12 critical peak events during each year of the study.

Control/information technology treatments include the deployment of IHDs. SMUD is offering IHDs to all opt-out customers in any given treatment group and to more than half of the opt-in customers in the treatment group. All participating customers receive web portal access, customer support and a variety of education materials.

Experimental Design

Due to the variety of treatments, the study includes three different experimental designs: randomized controlled trial (RCT) with delayed treatment for the control group, randomized encouragement design (RED) and within-subjects design (see Figure B-1).

In all three cases, AMI-enabled residential customers in SMUD's service territory are initially screened for eligibility and then randomly assigned to one of the seven treatments or the RED control group.

For the two treatments that are included in the RCT "Recruit and Delay" study design, customers receive an invitation to opt in to the study where participating customers receive an offer for a specific treatment. Upon agreeing to join the study, customers are told if they are to begin receiving the rate in the first year of the study (i.e., June 2012) or in the summer after the study is complete (i.e., June 2014).

For two of the three treatments that are included in the RED, customers are told that they have been assigned to a specific identified treatment but have the ability to opt out of this offer. Those who do not opt out receive the indicated treatment for the duration of the study. Those who opt out are nonetheless included in the study's evaluation effort but do not receive the indicated treatment. For one of the three RED treatments, customers receive an invitation to opt in to the study where participating customers receive a specific treatment. Customers that opt in are then assigned to receive the treatment in year 1 of the study (i.e., 2012).

For the two treatments that are included in the within-subject design, customers are told they have been assigned to either the Block w/CPP treatment or the TOU w/CPP treatment with technology.[e] In the former case, customers only have the ability to opt in to this specific treatment. In the latter case, customers only have the ability to opt out of this specific treatment.

---

[e] The within-subjects method was designed to use no explicit control group; instead it estimates the effects of the treatment for each participant individually, using observed electricity consumption behavior both before and after becoming a participant in the study as well as on critical peak event and non-event days. However, the control group selected for the RED design may be used as a control group.

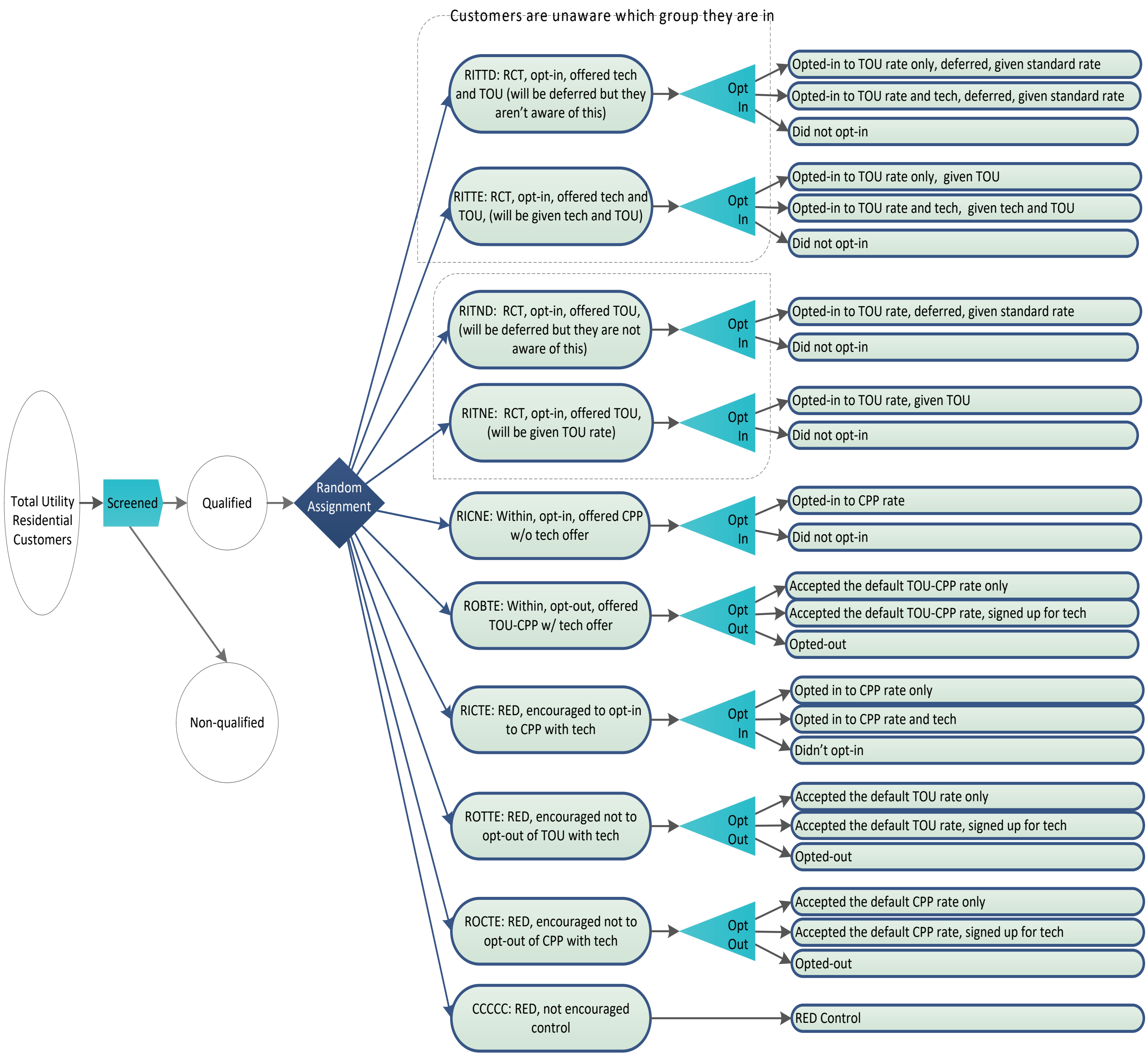


Figure B-1. SMUD Recruitment Process

Validation of Randomization

Table 1 provides summary statistics by experimental group. The top three rows summarize information on daily consumption, the ratio of peak to off-peak energy consumption, and billing from the pre-treatment summer (June to September 2011). Sample households consume slightly less electricity than the average U.S. household - approximately 27 kWh per day during the four summer months compared to almost 31 kWh per day across the U.S. in 2011. The ratio of peak to off-peak usage provides one indication of a customer's exposure to the higher peak prices under CPP or TOU, and bill amounts reflect the average monthly bill in the pre-treatment summer. Bills in our sample are very close to the national average, reflecting that SMUD customers pay higher prices than the average U.S. residential customer. For all three variables, we also report t-statistics on the test that the mean for each treatment group equals the mean for the control group.[f] The t-statistic exceeds one for only one of these comparisons, suggesting that the randomization yielded groups with very similar means across these three variables.[g]

The ``structural winner'' variables measure the share of households that would pay less on either the CPP or TOU pricing policy, assuming no change in their consumption. (Following industry convention, we refer to households who would pay less as ``structural winners.'') Approximately half of all customers are estimated to be structural winners, based on consumption data collected before the intervention.[h]

---

[f] We also run t-tests comparing the opt-in to opt-out treatment groups, and find no statistically significant differences.

[g] Given that we will be analyzing consumption across hours of the day, we are particularly concerned about balance in consumption profiles. In addition to the ratio of peak to off-peak usage, Figure A.1 and A.2 provides a breakdown of consumption across all 24 hours of the day. Again, all four treatment groups look very similar to the control group.

[h] To identify structural winners under the CPP program, we simulate 12 CPP days in the pre-treatment period by choosing the 12 hottest non-holiday summer weekdays. To ensure that our estimates of structural winnership do not result from idiosyncratic variation on these 12 days, we also estimate specifications where we randomly select 12 of the 24 hottest non-holiday summer weekdays and recompute our estimate of pre-period CPP bills. We repeat this exercise 10,000 times and then average over the estimated pre-period CPP bills to obtain an alternative measure of structural winnership. The correlation between the two measures is 0.97.

Table 1: Comparison of means by treatment assignment

	Control group	Treatment groups CPP		TOU	
		Opt-in	Opt-out	Opt-in	Opt-out
Daily usage (kWh)	26.63	26.81	26.92	26.49	26.38
		(-0.82)	(-0.45)	(0.83)	(0.71)
Peak to off-peak ratio	1.77	1.77	1.78	1.78	1.78
		(0.02)	(-0.50)	(-0.57)	(-0.37)
Bill amount ($)	109.10	109.44	109.12	108.20	107.86
		(-0.34)	(-0.01)	(1.08)	(0.69)
Structural winner (CPP)	0.51	0.51	0.52	0.51	0.50
		(-0.51)	(-0.39)	(-0.11)	(0.70)
Structural winner (TOU)	0.34	0.34	0.35	0.34	0.33
		(-0.13)	(-0.15)	(0.41)	(1.14)
Households	45,839	9,190	846	12,735	2,407

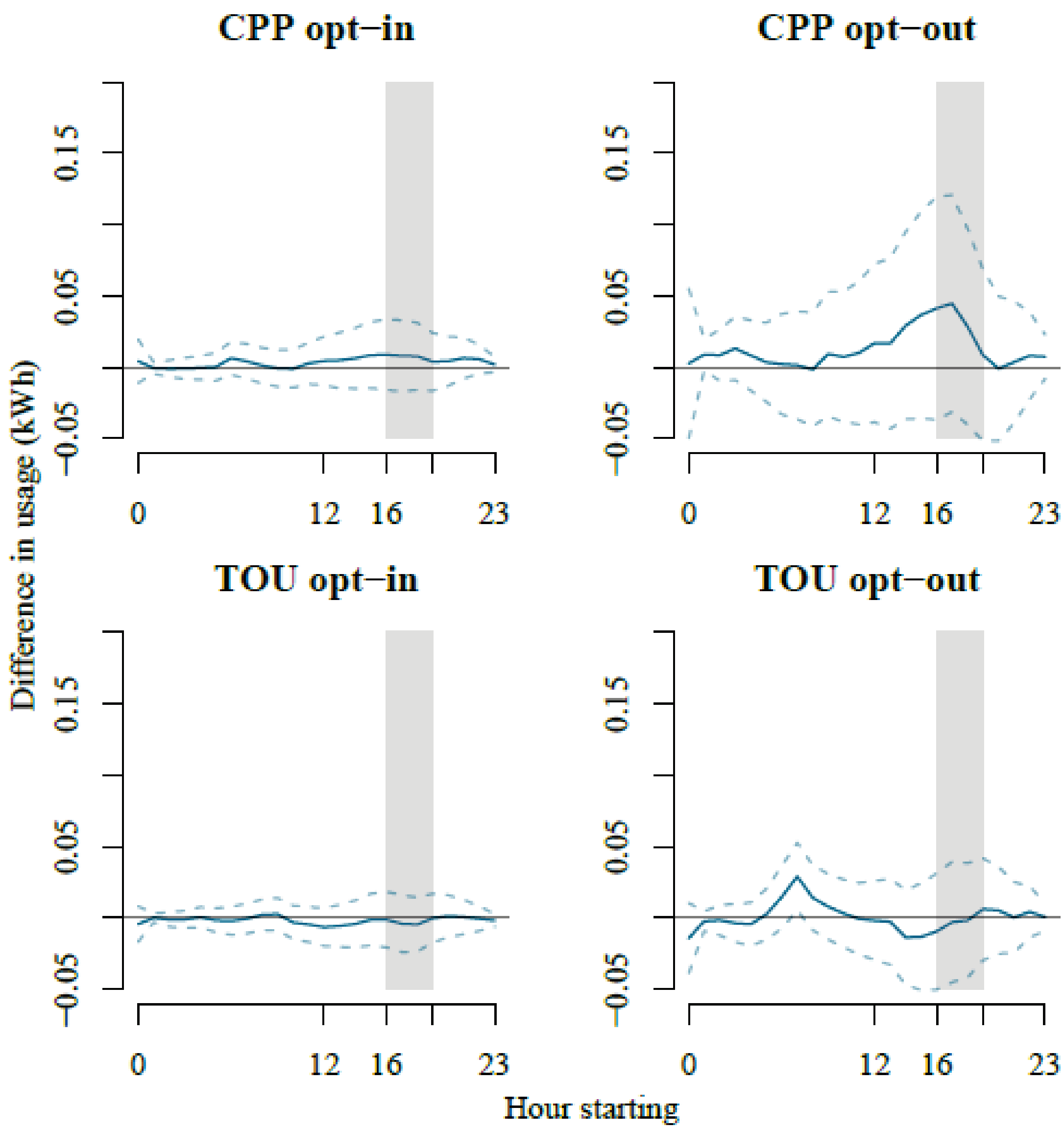
Notes: This table compares pre-period usage statistics across control and treatment groups. Cells contain group means and t-statistics (in parentheses) obtained from a two-sample t-test comparing means in the control group to means in the given treatment group. Daily usage is the average per-customer electricity usage over the pre-period summer. Peak to off-peak ratio is the average hourly consumption during peak periods (4-7pm on weekdays) divided by the hourly kWh used during non-peak times over the pre-period summer. Bill amounts reflect monthly bills over the pre-period summer. Structural winner is an indicator variable for whether the household would have experienced reduced bills in the pre-period summer had they been enrolled in either the CPP or TOU pricing plans.

Load Shape Balance across Treatment Groups

Because we analyze consumption across hours of the day, we are also concerned about balance in hourly consumption profiles. Figure A.1 and A.2 plots each treatment group's hourly electricity consumption overlaid with control group consumption, obtained from a regression of electricity consumption on a set of indicator variables for each hour. The left side of the figure compares customers who were offered the opportunity to opt-in to either the CPP or TOU treatment to control customers, while the right side compares customers who were defaulted on to either the CPP or TOU plan to the same control customers. The graph highlights the variation in electricity consumption over the day, from a low below .75 kWh in the middle of the night to a peak nearly three times as high at 5PM. This consumption profile is typical across electricity consumers around the country, although SMUD customers' peak consumption tends to be slightly later than for customers of other utilities.

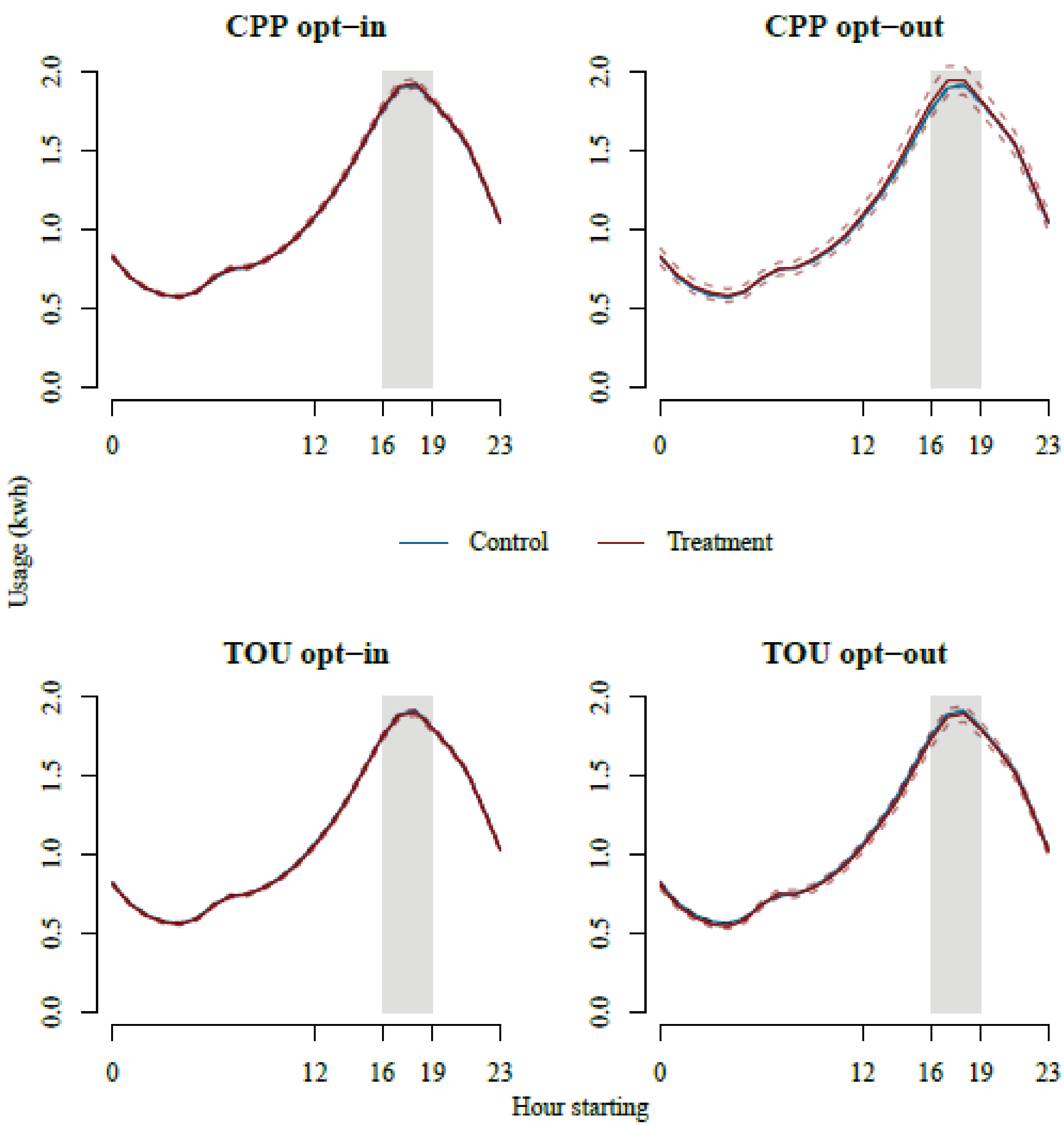
The graph also highlights that we cannot reject that both sets of treated households had statistically identical consumption profiles to the control households. The graphs in Figure A.1 and A.2 show the differences between treated and control, highlighting that these are well within the 95 percent confidence intervals for all hours. The standard errors for the CPP opt-out group are notably larger since that group had one tenth as many households.

Figure A.2: Difference between treatment and control groups' electricity consumption prior to treatment



Notes: Figure depicts average difference in pre-treatment weekday electricity usage in kW between treatment and control groups. Lines represent regression coefficients from interactions between hourly indicator variables and a treatment indicator. Dashed lines indicate 95% confidence intervals, clustered by household. Vertical bars indicate peak period, between 4pm and 7pm.

Figure A.1: Pre-treatment electricity usage



Notes: Figure depicts average pre-treatment weekday electricity usage in kW. Panels plot average treatment group hourly electricity consumption overlaid with control group consumption, with coefficients and standard errors clustered by household obtained from a regression of electricity consumption on a set of indicator variables for each hour. Dashed lines indicate 95% confidence intervals.

Appendix B. C-tree Algorithm

The ctree algorithm for recursive partitioning identifies splits in explanatory variable values (i.e. segments customers into groups with values greater than or less than a specific value conditional on all preceding splits) that provide the most statistically significant difference in the explained variable across each group. In our case the explanatory variables are the energy use variables listed in Table 1 and the explained variable is the enrollment choice.

The c-tree stops branching when correlation between the input variables and the response cannot be established with sufficient confidence (This is based on the failure to reject the null hypothesis that the input variables and response variable are independent. Our criteria for rejection of the null hypothesis is less than 95% confidence).

The leaf nodes produced by the c-tree algorithm can be thought of as algorithmically generated customer segments that are defined by the metrics of the branching nodes above them. To confirm the resulting groups used in the paper indeed have heterogeneous enrollment probabilities, we first conduct a pair wise hypothesis testing among the groups as follows.

$\boldsymbol{H_0}$: $p_i = p_j$

$\boldsymbol{H_A}$: $p_i \neq p_j$

Where p_i and p_j are the enrollment probability of group i and j, with the estimated value of

$\hat{p}_i = \frac{n_i}{N_i}$, n_i is the number of enrolled customers in group i and N_i is the total customer in group i). The test statistic is computed as

$$z = \frac{\hat{p}_i - \hat{p}_j}{\sqrt{(\hat{p}(1-\hat{p})(\frac{1}{N_i} + \frac{1}{N_j})}}$$

Where $\hat{p} = \frac{N_i\hat{p}_i + N_j\hat{p}_j}{N_i + N_j}$.

We conduct two-tail tests at 95% confidence interval for all possible pairs. The pairs are considered significantly different when the absolute value of test statistic (z) is greater than 1.96. As we can see in Table 2 that 29 out of the 36 pairs (i.e. 81%) are significantly different at the 95% confidence interval.

Table 2. Values of z statistic of the enrollment difference between pairs of c-tree groups.

	F	I	C	G	B	A	H	E	D
F		4.600	-3.70	1.20	-4.90	-6.90	3.200	0.00	-2.90
I	4.600		-7.40	-3.20	-8.40	-10.00	-1.00	-4.70	-7.00
C	-3.70	-7.40		4.300	-1.10	-2.50	5.900	3.800	1.30
G	1.20	-3.20	4.300		-5.30	-6.90	2.000	-1.20	-3.60
B	-4.90	-8.40	-1.10	-5.30		-1.20	6.800	5.000	2.600
A	-6.90	-10.00	-2.50	-6.90	-1.20		8.200	7.200	4.500
H	3.200	-1.00	5.900	2.000	6.800	8.200		-3.20	-5.30
E	0.00	-4.70	3.800	-1.20	5.000	7.200	-3.20		-3.00
D	-2.90	-7.00	1.30	-3.60	2.600	4.500	-5.30	-3.00	

Another way to define the performance of the algorithm at the task of predicting heterogeneous enrollment is to quantify the degree to which the enrollment rates of the customer segments deviate from the mean enrollment rate of the full sample. Since the sample mean is unaltered by grouping customers, every segment with higher than average enrollment will be offset by one or more segments with lower enrollment. We define a "spread score," θ, as the customer weighted mean of the absolute value of the segment deviations from the mean, where *S* is the group of all segments, i.e. leaf nodes, defined by the tree, n is the total number of customers, n_s is the number of customers in each segment, s, p is the sample wide enrollment rate and p_s is the enrollment rate for segment s.

Eqn. 3

$$\theta = \frac{1}{n}\sum_{s \in S} n_s |p - p_s|$$

The higher the spread score, the better a given tree has done at distinguishing between segments with low propensity to enroll and high propensity to enroll. We note that this score can be calculated for either in sample fits, or out of sample predictions. In our case, the spread score is the average absolute deviation in enrollment rate for each customer, so a θ value of 0.02 equates to an average expected deviation of 2% from the sample mean enrollment.

To evaluate the robustness of the spread score coming out of c-tree, we created a random forest of 250 c-trees, with each fit based on metrics from a random sub-sample of just over 60% of customers. We then computed the spread scores for each of the resulting trees to see the variation in goodness of fit created by the random sampling of training data. Finally, we computed 100 iterations of spread scores for random subsets of out-of-sample customers for each of the top 10 trees. These tests were used to confirm that the c-tree segments could be used to pre-identify significant numbers of customers who would turn out to be significantly more likely to enroll in the TOU pricing program than average.

The distribution spread scores across the random forest of 250 ctrees illustrates a wide range in scores but even the lowest scores are significantly higher than zero. This suggests that all examined trees have predictive power and that the highest performing ones can isolate large samples of customers with enrollment rates averaging 2-2.5% higher than the sample mean. The red line in Figure 1 indicates the spread score of the tree used in our paper to segment customers. We can see its performance is better than average but within a normal range defined by the performance of individual trees in the forest.

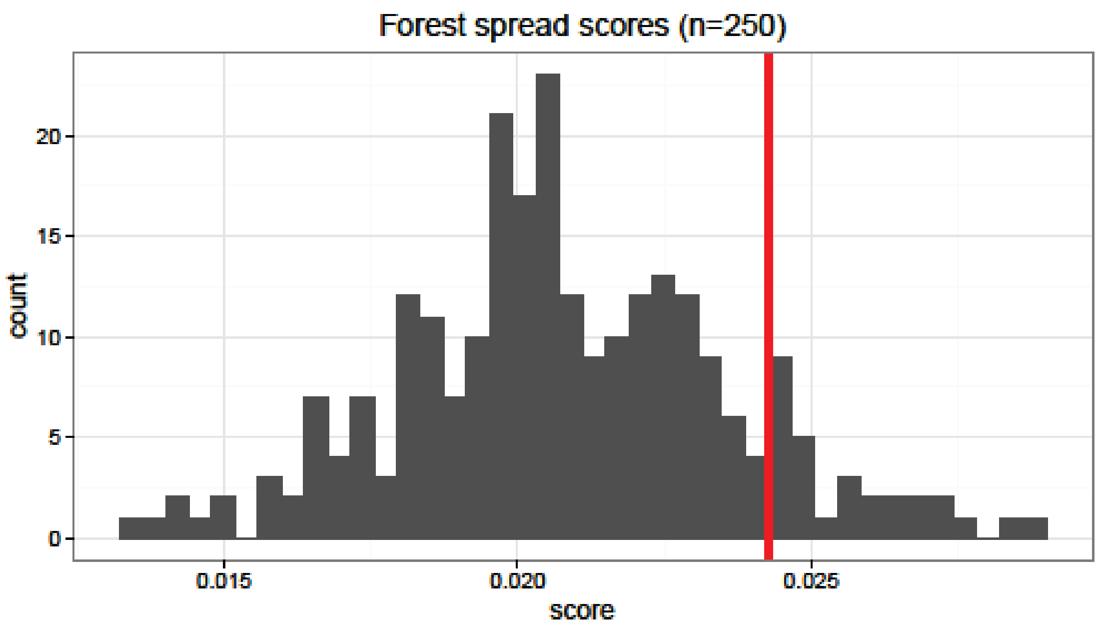


Figure 2: Spread scores from the 250 trees in the random forest. Red line indicates spread score for the tree used in our paper that derived from the full sample to segment customers

Appendix C. Deriving the Energy Characteristic Metrics

Here we provide somewhat more detail on how we derive the entropy variable. The entropy variable is meant to characterize overall variability in daily household consumption patterns and is generated in the following way. We first derive representative daily baseload usage load shapes with a clustering analysis technique building from the adaptive k-means approach used by Kwac, Flora and Rajagopal [1]. We introduce two innovations building off of this previous approach: first, we remove the daily minimum usage (as a proxy for base load) from the load profiles before normalizing the daily load shapes and applying the clustering algorithm, and second we applying an iterative truncation procedure that removes cluster centers with low member counts. We then reassign the load shapes removed from these eliminated outlier clusters based on their best fit among the remaining clusters. To accomplish this iterative truncation we define a threshold V to be the acceptable level of the "violation rate," which is the fraction of load shapes violating the original adaptive k-means θ threshold condition. The truncation algorithm repeats until the violation rate equals V. The truncation algorithm proceeds as follows: (1) identifies the smallest clusters whose members comprise the fraction V of the total number of load shapes; (2) removes these clusters; (3) reassigns the load shapes that were members of the removed clusters into the remaining clusters; (4) computes the violation rate, which is the fraction of the load shapes with relative squared error (RSE; defined in [1]) > θ; and (5) repeats steps 1-4 as long as the computed violation rate is less than the user-specified threshold, V. Once this threshold is reached, the procedure stops.

We subtract the minimum daily usage from each daily load profile in order isolate baseload usage from more variable discretionary usage. We employ the iterative truncation step in order to focus on cluster centers that are most representative of the majority of usage patterns, rather than outlier patterns.

Clustering of these daily discretionary usage patterns is based on hourly consumption data from the pre-treatment year. The clustering algorithm groups load profiles into 99 unique baseload load shape clusters. Entropy in this load shape cluster assignment, as a metric of variability, is calculated as the degree to which a given household's pre-treatment daily load profile assignment varies across these 99 representative load shape clusters across all days. We do this in the following way: first, as shown in Equation 1, for each household n, we calculate the relative occurrence frequency of each representative load shape cluster, $p_n(L_i)$ that is assigned to that household's daily series. Here T is the total number of days (indexed by t) in the time series of household n, and $I^i(t)$ is an indicator equal to one if daily load shape t is assigned to representative load shape cluster i, zero otherwise.

Eqn. 1

$$p_n(L_i) = \frac{\sum_{t=1}^{T} I^i(t)}{T}$$

Then the entropy of household n is computed as follows in Equation 2, where, i continues to be the index over the 99 representative load shape clusters (L).

Eqn. 2

$$entropy_n = -\sum_{i=1}^{99} p_n(L_i) \ln(p_n(L_i))$$

References for Appendices

[1] J. Kwac, J. Flora, R. Rajagopal, Household energy consumption segmentation using hourly data, IEEE Transactions on Smart Grid 5(1) (2014) 420-430.