(Multi-) Leveling Up Savings Estimates: Using Hierarchical Models to Optimize a Home Energy Report Program

Seth Wayland, Opinion Dynamics Corporation
Olivia Patterson, Opinion Dynamics Corporation
Kara Downey, Opinion Dynamics Corporation

ABSTRACT

As home energy reports reach more residential customers, it is critical to understand which customer segments are realizing the greatest energy savings. How can groups of customers with high energy savings be identified? How can we cost-effectively target customers for program expansion? Does exposure to a home energy report lead to additional savings behavior? The answers to these questions are invaluable in helping utilities reach beyond the "low-hanging fruit" and continue to expand energy savings as programs mature.

Traditional evaluations of RCT-style home energy report programs typically account for individual variance in responses to treatment by including a household-specific intercept in regression models(fixed effects).While this method creates an accurate estimate of overall savings by controlling for all unobserved household-specific characteristics, it cannot effectively estimate household-specific treatment effects. If individual savings estimates are calculated at all, it is typically via household-specific regressions. Multi-level models combine these two approaches for the best of both worlds; correcting for individual variance and yielding household-specific savings estimates that will still be correct even in the face of correlated observations.

Our paper presents the results of a multilevel approach to calculating individual savings estimates for a multi-year home energy savings report program. Multi-level models allow us to generate individual-level savings estimates that take household and demographic characteristics, as well as traditional controls like weather into account. These individual-level estimates allow us to identify groups of high, low, and negative savers, and to investigate whether participants tend to stay in the same group or move into a higher or lower savings group over time.

Introduction

Home energy reports (HER) are an increasingly common part of the energy efficiency landscape. These programs are usually implemented on a large scale, with tens (or hundreds) of thousands of customers receiving energy usage reports that attempt to motivate them to make small changes to reduce consumption. Evaluators usually assess impacts by estimating average savings per report across the entire program. While average savings provide a consistent and accurate result for claiming savings, they provide limited value in terms of optimizing program design and implementation. What if, in addition to providing an overall savings number, evaluators could provide insight into which customers were responding most positively and most negatively to the reports?

In this study, we use multilevel modeling to show that overall savings numbers hide a great deal of variation in program savings at the household level. Specifically:

- Certain customers achieve significantly higher savings than others
- Some customers (~40% gas and electric) actually experience negative savings
- Customers who are negative savers in the first year rarely evolve into positive savers over time.

Taken together, these findings suggest that utilities and implementers can increase HER savings by moving away from a one-size-fits-all approach, expanding outreach to customers likely to be high savers, and changing offerings for low and negative savers. Further research is needed to determine whether and how one can successfully target members of different savings groups. We are now conducting surveys to better understand these groups and optimize future messaging and delivery efforts.

In the next section of this paper, we will give a general overview of HER programs. Next, we will introduce multilevel modeling and explain how we applied it to a HER billing analysis. After that, we will discuss the different savings groups that we identified. Finally, we will close with implications for future research and development of HER programs.

Home Energy Report Program Background

HERs are the most common type of a wider class of energy efficiency interventions known as behavioral modification programs (Patterson 2014). Such programs achieve savings by changing customer usage habits as well as technology choices. HER programs typically rest on two key drivers of behavioral change: historical usage and social norming. The program logic argues that giving customers information about their own past usage and about that of their peers will change their beliefs about “normal” energy consumption and create social pressure to use less. HERs leverage both billing data and publically available data to provide energy efficiency information, usage history, and benchmarking to participants. The reports are typically provided by a third party (e.g. Aclara, Tendril Energy, C3 Energy, Opower).

There is strong evidence to suggest that these programs produce energy savings. Most HER programs are evaluated via statistical analysis of billing records of participants that are compared to those of control or comparison group customers. Some HER reports are distributed within a randomized control trial framework, while other programs must be evaluated in a quasi-experimental manner. Regardless of the specific evaluation method, most HER program evaluations find evidence of small but consistent household-level savings. When these small savings are applied to a large number of customers, overall program savings can be substantial.

However, these evaluations rarely probe more deeply into whether and to what degree individual households respond differently to reports. The typical statistical methodology for calculating program savings (discussed in detail below) is designed to produce an overall average savings value across all participating households, not to estimate individual savings. Some evaluations use surveys to attempt to capture differences in energy behavior before and after the start of the program and across customers. However, these surveys have met with mixed success, partially due to the challenges of gaining enough responses and reducing the influence of faulty recall, social desirability bias, and measurement errors.

In order to deepen savings (not to mention improve customer engagement and satisfaction), HER programs need to look beyond average values and understand the diverse and sometimes surprising individual patterns that create the overall results. The data and technology needed for building more customized HER programs exists. In recent years, HER providers have increasingly leveraged “big data” to provide more tailored reports, deployed email reports, web portals, and mobile apps in addition to traditional paper reports, and have expanded motivational techniques to include goal setting, competitions, and using customers' online social networks (Opinion Dynamics and Navigant 2013). What is needed is a better way of understanding which customers are responding well to current efforts, and which customers might benefit from a modified approach. Multilevel modeling, described in detail in the next section, provides one way of looking "under the hood" of the overall results and identifying household-level savings patterns.

Methodology

Multilevel modeling has become the gold standard for social science researchers interested in studying outcomes that are affected by both individual- and group-level variables. For example, performance on standardized tests is often modeled as a function of both personal characteristics such as family socioeconomic status, age, or ethnicity *and* school-level characteristics such as the student-teacher ratio, funding level, and style of instruction. Similarly, one could think of savings from an HER program as a function of both program-level factors such as weather in the utility service territory or the design of reports *and* household-level factors such as type of house or the level of competitiveness of the person reading the report.

The term "multilevel modeling" refers to the fact that certain coefficients in the regression are themselves modeled by another, "higher-level" equation. This is advantageous because it allows the model to take both individual- and group-level variance into account when estimating coefficients. We used Equation 1 to estimate household-specific changes in energy consumption for the treatment group in the post-period using utility billing data. All of the calculations and modeling used R statistical software, with multilevel models using the lme4 package.

Equation 1. Multilevel Model

$$ADC_{it} \sim N(\alpha_i + \theta_i Treatment_t + \beta_1 HDD_t + \beta_2 CDD_t + \beta_3 PreADC_i + \beta_4 PreADC * Treatment_{it}, \sigma^2_{ADC}),$$

$$for\ t = 1, \ldots, t, i = 1, \ldots, n_j$$

$$\begin{pmatrix} \alpha_i \\ \theta_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_\alpha \\ \mu_\theta \end{pmatrix}, \begin{pmatrix} \sigma^2_\alpha & \rho\sigma_\alpha\sigma_\theta \\ \rho\sigma_\alpha\sigma_\theta & \sigma^2_\theta \end{pmatrix} \right), for\ i = 1, \ldots, n$$

Where:

ADC_{it}= Average daily consumption (kWh or therms) for household i at time t

α_i= Household-specific intercept for household i

θ_i= Household-specific change in consumption for the treatment group in the post period

β_1= Coefficient for HDD (Heating Degree Days)

β_2= Coefficient for CDD (Cooling Degree Days)

β_3= Coefficient for PreADC (Pre-Period average daily consumption)

β_4= Coefficient for PreADC by Treatment interaction

σ^2_{ADC}= Variance of ADC (average daily consumption)

μ_α= Mean of household-specific intercept

μ_θ= Mean of household-specific change in consumption due to treatment

σ^2_α= Variance of household-specific intercept

σ^2_θ= Variance of household-specific change in consumption due to treatment

$\rho\sigma_\alpha\sigma_\theta$= Covariance of household-specific intercept and change in consumption

We drew data for this analysis from several sources, including program-tracking data, customer billing data, and demographic and household data purchased from Experian. The billing data includes monthly records for over 250,000 electric, gas, and dual-fuel customers. The duration of exposure to the program varies: the program we analyzed is entering its seventh year, and new cohorts are added to the program each year.

After estimating the individual savings for each customer, we used those individual savings estimates to group customers into five categories (high, medium, neutral, negative and very negative savers) and analyzed the correlation of these categories with demographics and household characteristics drawn from third-party data. We repeated this analysis on three years of program data to analyze changes in savings group membership over time. We will discuss these results in detail in the next section. Before doing that, we will briefly touch on the advantages and disadvantages of using multilevel modeling to obtain these results.

Advantages of Multilevel Modeling. Multilevel modeling results provide insights into customer behavior that cannot be achieved using the tools that evaluators usually employ for billing analysis. Most traditional billing analyses use either linear or lagged dependent variable regression with household-level fixed effects. These models generate household-specific intercepts and precise estimates of overall savings, but do not provide household-specific savings estimates. As a result, it is impossible to tell whether all customers are affected by the program in approximately the same way, or if some customers achieve savings significantly higher or lower than average customer savings.

Figure 1 below illustrates the differences between the results than can be estimated by a typical fixed effects model (left panel) and a multilevel model (right panel) using hypothetical data. The X-axis represents number of days in an HER program, and the Y-axis represents total program savings. The thick dashed lines represent average program savings, and the solid lines represent the savings of individual customers. It is easy to see that the overall estimates of program savings, which take all of the participants into account, are quite similar. However, the two types of models make very different assumptions about underlying customer behavior. In the fixed effects model, customer-level differences do play a role in the form of the different intercepts. However, each hypothetical customer gains the same amount of savings from being in the program for an additional day (i.e., the slopes of the lines are the same), and one would assume that maximum program savings would be achieved by adding as many customers as possible. In the multilevel model, customers respond to treatment in different ways. Customer C achieves huge savings, Customer A resembles the overall average, Customer C barely saves anything, and Customer D actually has negative savings. Overall program savings would be increased by attracting more customers like Customers A and C and changing or perhaps even ending reports for customers like Customer B and Customer D.

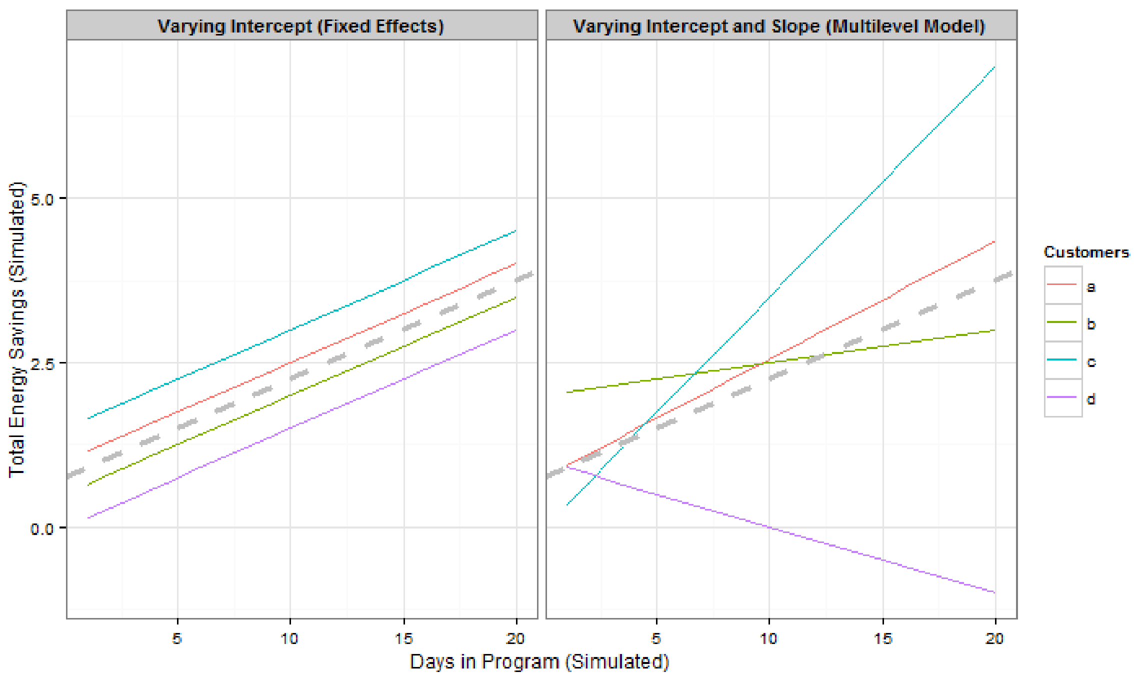

Figure 1. Comparison of Fixed Effects and Multilevel Models.

Numerous process evaluations and interviews with program participants and implementers suggest that customers respond differently to HERs, and that the real world looks more like the right-hand than the left-hand panel of Figure 1. It is also possible to estimate savings levels for individual households by running individual regression models for each participant. However, the multilevel modeling approach provides several important advantages over individual regression in establishing individual household savings levels. First, multilevel modeling statistically controls for weather differences between pre- and post-periods for an individual household as well as across households. In contrast, individual models solely control for weather differences between pre- and post-periods for an individual household. Second, multilevel modeling allows for modeling the influence of variables that do not change over time that apply to customers and for generating appropriate standard errors and statistical tests. Third, results from multilevel regression models adjust individual savings estimates based on control group usage during the treatment period, so the savings estimates are much closer to actual savings than results from individual regressions. Finally, information is shared across customers in multilevel models, so the unexplained variance in individual savings across participants is much lower when we make estimates using a multilevel model.

Another major advantage to multilevel modeling is its ability to estimate how much of the variation in savings is accounted for at each "level" of customer data. For example, a basic multilevel model that looked at the customer, neighborhood, and substation levels would indicate how much savings differed at each of those levels. This information would be extremely helpful in program design, as it would help implementers better understand whether to target different interventions at specific households or entire neighborhoods.

In short, multilevel modeling provides a useful compromise between estimating aggregate results but ignoring individual differences (traditional fixed effects models) and estimating individual results but overlooking program-level dynamics (individual regression models).

Limitations of Multilevel Modeling. Although it has many advantages, multilevel modeling, like any methodological tool, is not without its own shortcomings. First, it is computationally expensive. Many behavioral modification programs operate on a massive scale, often with hundreds of thousands of participants. Although computing power continues to improve rapidly, estimating hundreds of thousands of unique slopes and intercepts is not a trivial task. Estimation becomes even harder if, instead of comparing customer demographics after the fact, researchers include them in the customer-level equation in the model. Because these models are so complex to run, evaluators may find that they are constrained in the number of parameters that they can include relative to fixed effects models.

Second and closely related to the limitations above, the overall savings estimates produced by our multilevel model do not fully agree with the estimates produced by the various fixed effects models. This is because the variables in the multilevel model were chosen to maximize our ability to understand differences in household-level responses to treatment rather than to estimate overall savings with the highest possible precision. Fitting a multilevel model also requires much larger amounts of data than fitting a fixed effects model. At present, multilevel modeling is best understood as a potentially useful addition to traditional methods of billing analysis, not a substitute for them. The primary value of multilevel modeling lies more in its potential to improve program targeting and design than in calculating overall savings estimates, though future improvements in model computation could make multilevel models a good choice for overall savings estimates as well.

Third, and most importantly, the method is unproven in terms of its ultimate relevance to utilities. Our assumption that the different savings groups identified by the multilevel model would be useful targets for new forms of messaging has yet to be tested. Rolling out different reports to different savings groups would also require building a predictive model capable of accurately identifying different savings group members using data typically available to utilities or implementers. However, the potential benefits of improving one-size-fits-all HER programs make tackling these challenges worthwhile. More customized programs might also increase customer engagement and satisfaction, and enhance cost-effectiveness by increasing potential savings while decreasing the overall number of reports.

Results

We first examined savings for the most recent year of the HER program. The household-specific savings estimates obtained by the multilevel model showed not only that customers varied significantly in terms of the amount of energy they saved, but also that approximately 40% of customers (both gas and electric) actually had negative savings associated with the program. This is an important finding in its own right, as it suggests that residential HER programs could boost savings by enrolling more customers likely to benefit highly from the program and modifying their approach to negative savers.

Next, we split the participants into five savings groups based on their results from the multilevel model in order to explore the characteristics of high versus low savers. We created separate groups for gas and electric customers, so the same customer could be a high saver for gas and a low saver for electricity. Figures 2 and 3 below show the distribution of individual electric and gas savings elements with colors that indicate savings groups.

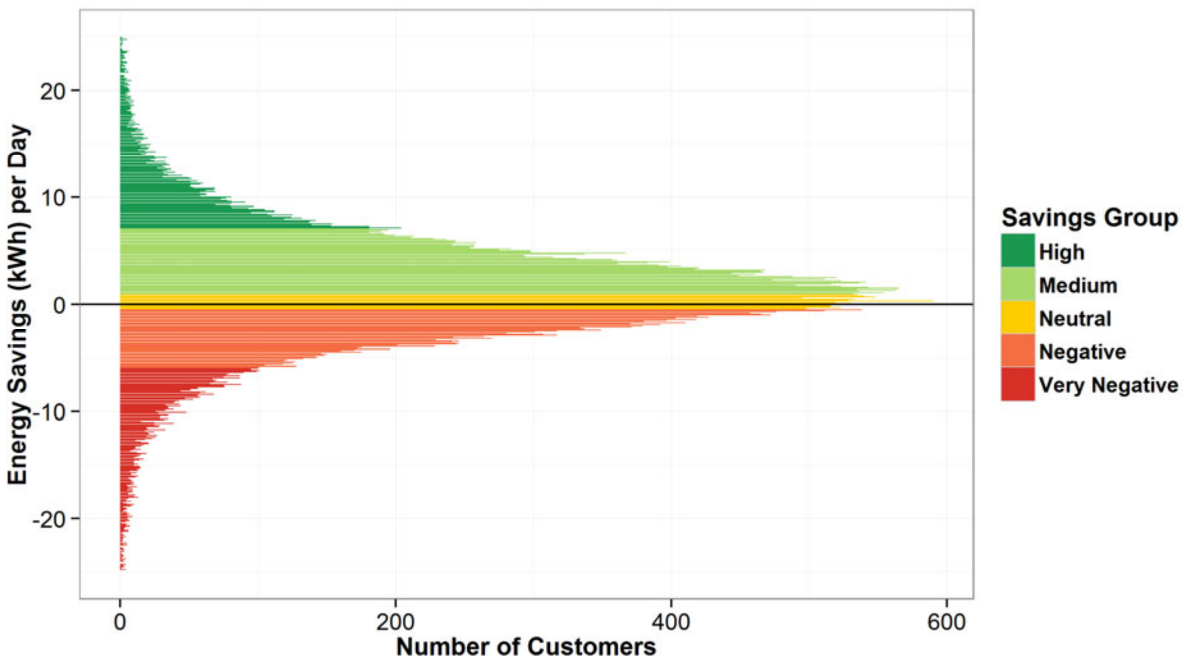

Figure 2. Distribution of individual electric savings estimates. Group cut-offs are as follows: High > 7 kWh; Medium > 1 & ≤ 7 kWh; Neutral > -0.5 & ≤1 kWh; Negative > - 6 & ≤ -0.5 kWh; Very Negative < -6 kWh

The variation in individual savings estimates is striking. High electricity savers save an average of 12.33 kWh per day, but some save over 20 kWh per day. Gas participants might save or increase usage by over 1.5 therms per day. It is important to realize that differences in savings might not only be a hostile reaction to the reports. For example, a participant might have a baby, start working from home, or expand a house. We discuss the implications of this later in the paper. For now, it is sufficient to note that removing these negative savers from the program would increase overall program savings.

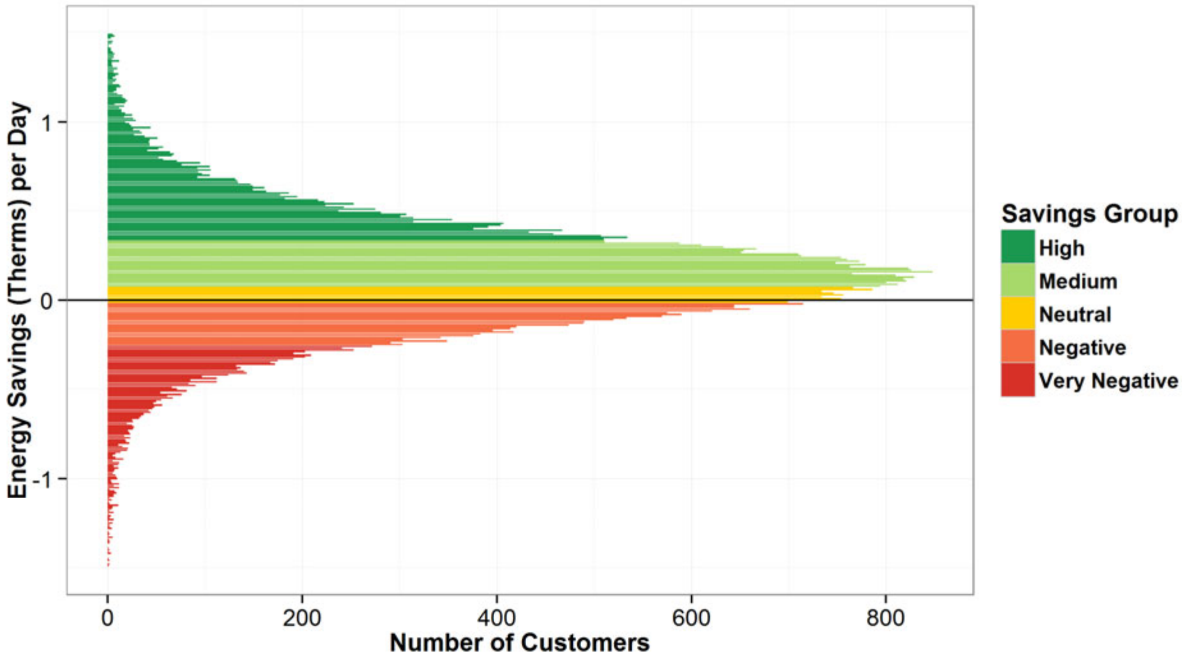

Figure 3. Distribution of individual gas savings estimates. Group cut-offs are as follows: High > 0.33 therms; Medium > 0.08 & ≤ 0.33 therms; Neutral > - 0.02 & ≤ 0.08 therms; Negative >-0.25 & ≤ -0.02 therms; Very Negative < - 0.25 therms

To put these numbers in context, overall savings range from 0.006 to 0.04 therms per day for gas customers and 0.15 to 0.6 kWh per day for electric customers in the traditional weather-adjusted fixed effects regression models depending on the cohort. Notice that these values are similar to the highest density portions of the individual savings distributions. The cohort-level averages produced by traditional models are accurate, but hide a great deal of household-level variation.

We next used utility data on past energy consumption and purchased demographic and housing data to better understand what might be driving differences in these savings groups. Table 1 describes each gas savings group in the most recent program year. The gas savings groups differ most on pre-period ADC (average daily consumption before the first report), age of building, and years of residence. On average, members of the highest savings group have the higher pre-period usage, fewer years of residence, and older homes than members of the other savings groups. They have higher winter ADC than all groups except the most negative savings group, a surprising finding that we will discuss in more detail later in the paper.

Table 2 shows the same measures as Table 1, but for electric savings groups. It is more difficult to discern demographic and household differences between the high savings group and other groups in this case, but the pre-treatment ADC differential is much larger between the negative and high savings groups than it is in the gas groups.

Table 1. Characteristics of Gas Savings Groups, Current Program Year

Savings Group	Percentage Savings	Average Therm Savings Per Day	Pre-ADC	Summer Pre-ADC	Winter Pre-ADC	Year Home Built	Years of Residence
High	22%	0.55	2.85	0.59	6.72	1967	7.8
Medium	9%	0.18	2.38	0.52	5.83	1972	9.2
Neutral	1%	0.03	2.30	0.50	5.76	1975	9.8
Negative	-6%	-0.11	2.19	0.47	5.66	1975	10.9
Very Negative	-21%	-0.44	2.61	0.51	6.93	1971	12.9

Table 2. Characteristics of Electric Savings Groups, Current Program Year

Savings Group	Percentage Savings	Average kWh Savings Per Day	Pre-ADC	Summer Pre-ADC	Winter Pre-ADC	Year Home Built	Years of Residence
High	28%	12.33	47.3	64.6	48.4	1972	10.0
Medium	10%	3.26	37.5	53.1	36.1	1973	9.7
Neutral	1%	0.23	30.6	44.2	29.1	1973	10.1
Negative	-10%	-2.38	27.6	39.0	27.3	1972	10.3
Very Negative	-37%	-11.44	36.7	49.2	39.6	1973	9.5

Consistent with other evaluations, the primary predictor of savings is pre-treatment usage. Higher users have a higher potential to save, and more often fall into the high saver group. We also found that housing characteristics and demographics are related to savings, though the magnitude of the relationship between the housing characteristics and savings varies by pre-treatment usage and interactions with other characteristics. To assess the importance of these non-linear relationships, we used a side effect of random forests modeling[1] that prioritizes the importance of the available variables for predicting savings (Liaw and Wiener 2002). The most predictive characteristics after pre-treatment usage were the age of the house, the customer's age, educational level, occupation, and number of people living at the residence. For gas, participants with older houses tend to save more, as do those who have lived in their home for less time. For electric, older participants, and those with fewer people living at the residence tend to save more.

Notably, some participants with relatively high usage fall into the very negative saver group in both the gas and electric analyses. This may mean that it could be difficult to select a group of customers with high propensity to save through choosing customers with high pre-treatment usage. It could be valuable from a program performance standpoint to adjust or stop

[1] Random forests makes many small recursive partitioning models with subsets of the variables, and uses the ordering of the partitions in the hundreds of models to order the predictive variables from most to least predictive.

delivery of the reports to very negative savers, or to target medium and high savers for future program waves. However, if such an approach is made, the implementation should use an experimental design to maintain design fidelity.

Temporal Evolution of Savings Groups. We next performed an analysis to see whether participants moved across savings groups over time. We examined the participant specific savings for the Original, Expansion 1, and Expansion 2 cohorts. To examine the evolution of savings groups, we included participants who stayed in the program for a minimum of three years, which makes the groups look slightly different than the PY7 participant specific groups above.

For this analysis, we expected that savings would increase from the first year of participation to the third, as participants are able to make more program related changes over time (Allcott and Rogers 2014). What we found was that for some participants this was the case, but some negative savers increased their usage more over the years of participation and moved from being negative savers to very negative savers.

Figure 3 and Figure 4 show the temporal evolution of the proportion of participants who fall into each savings group. Initially, nearly all participants fall into the middle three savings categories, and over time, some move into the extremes. We expected to see an increasing spread of savings over time with evolution of some customers from lower to higher savings as they made behavioral and equipment changes. The increase in the size of the very negative savings group may mean that some participants are responding to the home energy reports in ways that increase usage. Of course, it could also mean that people are enlarging their homes, expanding their families, or making other changes that cause their energy needs to grow. It could even be the case that such people are successfully incorporating tips from their HERs and using less energy than they otherwise would given the changes in their lives. This is another reason why our method, though promising, requires additional validation.

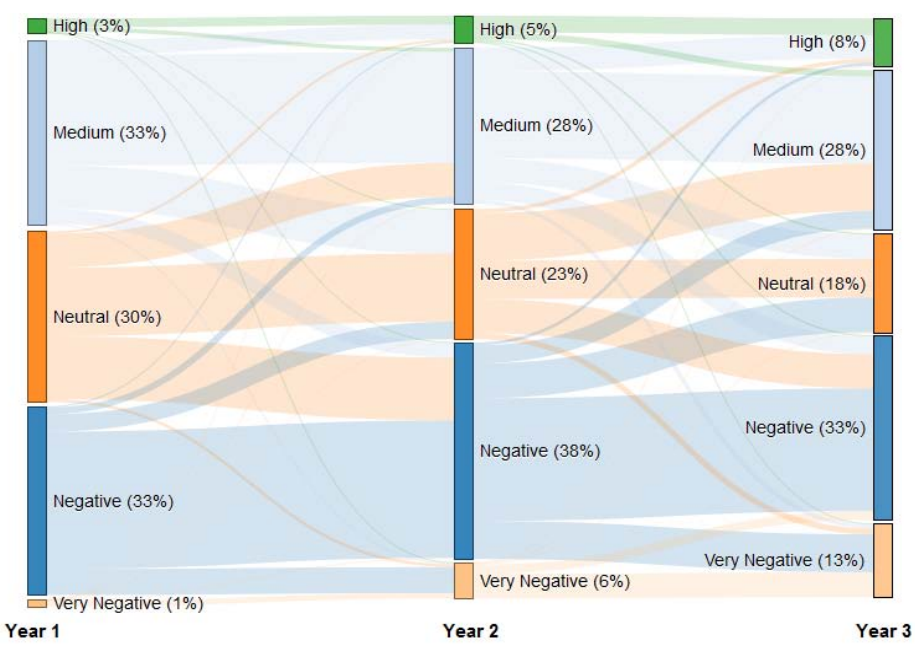

Figure 3. Gas Savings Group Evolution

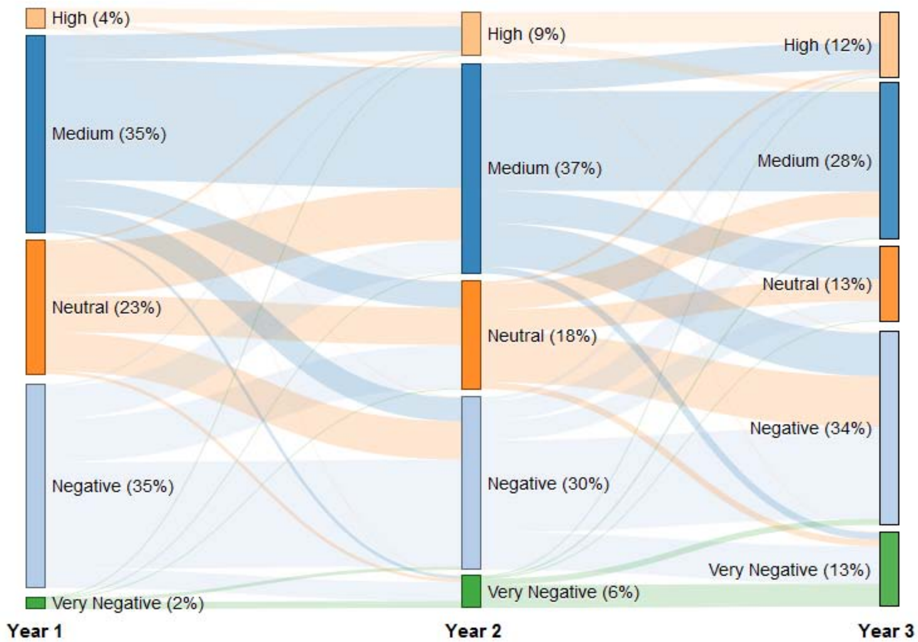

Figure 4. Electric Savings Group Evolution

We found that approximately 40% of customers stayed in the same savings group over the three years, while about 20% moved one group to higher savings and 20% moved one group to lower savings. Of the remaining 20% of customers, we found that slightly more customers moved more than one step toward lower savings than moved more than one step toward higher savings.

One interesting and potentially useful finding is that those customers who were in the very negative savings group in the first year very rarely achieved positive savings. For instance, in the gas analysis, 90% of those who started as very negative savers remained in the very negative or negative groups for all three years of the analysis. These customers might benefit from significant modifications to the reports they receive or from stopping reports entirely.

Conclusions and Implications for Behavioral Modification Programs

This study leveraged multilevel modeling, a technique from the social sciences, to demonstrate the diversity in participant responses to a residential HER program. This variation in responses suggests that there is considerable scope for these types of programs to increase savings by moving away from a one-size-fits-all approach to home energy reports and studying the characteristics of high, medium, and low savers in order to target program offerings more effectively. For example, a utility might recruit new cohorts mostly made up of customers similar to current medium and high savers, or stop sending reports to negative savers.

This study also raises important questions for future research and program design. First and most important is the need to understand why negative savers are negative. Is it a poor reaction to messaging? Is it the result of lifestyle changes, e.g. a participant who began to work at home or who expanded a house? These causes have different implications for program improvement. Randomly stopping reports to a subset of the participant group and starting them with a subset of the control group might be a way to begin to tease these differences out. Implementers could also consider A/B tests with different messaging strategies to negative savers.

Additionally, if customers are included or excluded based on their savings profile, utilities and implementers should take care to do this in a manner consistent with the program's randomized control trial framework in order to avoid compromising the integrity of the experimental design. Put concretely, from an evaluation standpoint it would be problematic for an implementer to have a treatment group composed only of high savers and a control group only of very negative savers. It is also important to remember that multilevel modeling is not a panacea. The large participant numbers typical of behavioral modification programs and the computational intensity of multilevel methods mean that, for now, researchers are limited in the types of models they can estimate. These models are more useful for disaggregating individual savings than for estimating overall program savings, and should not be viewed as a substitute for traditional fixed effects models.

Given rising energy efficiency standards (as well as rising temperatures and concern about global climate change), it is critical that utilities and implementers find new ways to increase program savings. Multilevel modeling, and the household-level savings estimates that it can generate, represents a promising way forward.

References

Allcott, H. and Rogers, T. 2014. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." *American Economic Review*. 104(10): 3003-37.

Liaw, A. and Wiener, M. 2002. "Classification and Regression by randomForest." *R News* 2(3): 18-22.

Opinion Dynamics with Navigant. 2013. *Massachusetts Cross-Cutting Behavioral Program Evaluation Integrated Report*. Prepared for Massachusetts Energy Efficiency Advisory Council and Behavioral Research Team. http://www.rieermc.ri.gov/documents/2013%20Evaluation%20Studies/ODC_2013_Cross_Cutting_Behavioral_Program_Evaluation.pdf

Patterson, O. 2014. "The Adolescent Years of Behavioral Programs: Optimizing Behavioral Program Design through Energy Savings Persistence." Presentation at the 2014 Behavior, Energy, & Climate Change Conference. http://beccconference.org/2014-presentations/