

NBER WORKING PAPER SERIES

HARNESSING POLICY COMPLEMENTARITIES TO CONSERVE ENERGY:  
EVIDENCE FROM A NATURAL FIELD EXPERIMENT

John A. List  
Robert D. Metcalfe  
Michael K. Price  
Florian Rundhammer

Working Paper 23355  
<http://www.nber.org/papers/w23355>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
April 2017

We would like to thank Marc Laitin for the opportunity to partner with Opower on this project and his excellent guidance throughout the project. We are also indebted to John Balz, Richard Caperton, Jim Kapsis, and many others at Opower for sharing data and offering insights. Opower provided the data analyzed in this paper to the authors under a nondisclosure agreement. The authors and Opower structured the agreement in a way that maintains the authors' independence. In particular, the agreement stipulates that Opower has the right to review the publication prior to public release solely for factual accuracy. Hunt Allcott, Eric Budish, Stefano DellaVigna, David Rapson, and participants of the 2017 ASSA meetings offered valuable comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by John A. List, Robert D. Metcalfe, Michael K. Price, and Florian Rundhammer. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Harnessing Policy Complementarities to Conserve Energy: Evidence from a Natural Field Experiment

John A. List, Robert D. Metcalfe, Michael K. Price, and Florian Rundhammer

NBER Working Paper No. 23355

April 2017

JEL No. C93,D03,Q4

**ABSTRACT**

The literature has shown the power of social norms to promote residential energy conservation, particularly among high usage users. This study uses a natural field experiment with nearly 200,000 US households to explore whether a financial rewards program can complement such approaches. We observe strong impacts of the program, particularly amongst low-usage and low-variance households, customers who typically are less responsive to normative messaging. Our data thus suggest important policy complementarities between behavioral and financial incentives: whereas non-pecuniary interventions disproportionately affect intense users, financial incentives are able to substantially affect the low-user, “sticky households.”

John A. List  
Department of Economics  
University of Chicago  
1126 East 59th  
Chicago, IL 60637  
and NBER  
jlist@uchicago.edu

Robert D. Metcalfe  
University of Chicago  
Saieh Hall for Economics  
5757 S. University Avenue  
Chicago IL, 60637  
rdmetc@gmail.com

Michael K. Price  
Experimental Economics Center  
Andrew Young School of Policy Studies  
Georgia State University  
P.O. Box 3992  
Atlanta, GA 30302-3992  
and NBER  
mprice25@gsu.edu

Florian Rundhammer  
The Department of Economics  
Andrew Young School of Policy Studies  
Georgia State University  
PO Box 3992  
Atlanta, GA 30302-3992  
florian.rundhammer@gmail.com

## I. INTRODUCTION

Behavioral economics has matured to the point where theorists are leveraging psychological insights to improve their models and policy-makers are drawing upon behavioral results to develop new strategies to influence decision-making. One particular result that has attracted increasing attention is the power of injunctive norms and social comparisons as a means to promote behavioral change. Social comparisons have been applied in a variety of settings, including voting participation (Gerber and Rogers, 2009), charitable giving (Frey and Meier, 2004; Croson and Shang, 2008; Shang and Croson, 2009), retirement savings decisions (Beshears et al., 2015), tax compliance (Hallsworth et al., 2017), and water conservation (Ferraro and Price, 2013; Brent et al., 2015). In this study, we focus on perhaps the most popular application of descriptive norms in the literature – energy conservation as driven by the receipt of Home Energy Reports (HERs) from Opower (see Allcott, 2011; 2015; Costa and Kahn, 2013; Ayres et al., 2013).<sup>1</sup>

Results from this literature suggest two stylized facts. First, households exposed to the HER reduce subsequent energy use significantly relative to a control group. In an important study spanning 111 distinct experimental sites across the US, Allcott (2015) identifies economically meaningful average treatment effects for *all* experiments. Yet, there are important heterogeneities, with most studies suggesting reductions in use that range from one to two percent. Second, the observed treatment effects are largely driven by high usage customers (Allcott 2011; Ferraro and Price, 2013). For example, Ferraro and Price (2013) find a fundamental difference in the effect of norm-based messages across low and high-use households – intensive use households experience treatment effects that are nearly double low use households.<sup>2</sup>

---

<sup>1</sup> The Home Energy Report includes information comparing a household’s energy use to that of a carefully chosen set of neighbors along with energy conservation tips designed to help customers understand ways to reduce energy use.

<sup>2</sup> Allcott (2011) provides similar evidence of heterogeneity across high and low-user groups. Although estimated treatment effects weakly increase with percentile of pre-intervention usage, the observed effects are statistically insignificant for households in the lowest deciles and exceed the two percent threshold for those in the highest deciles.

These stylized facts highlight the potential of behaviorally motivated policies, such as social comparisons, but leave open two important issues. First and foremost, can one design complementary strategies to move the needle and increase overall reductions in a quest to meet ambitious conservation goals? Second, can these complementary instruments affect choices of customers that are typically less responsive to social comparisons, i.e., lower user groups, without compromising the effect of the program for more responsive parts of the customer distribution? In this paper, we set forth to address these questions by presenting results from a natural field experiment conducted in partnership with Opower. The experiment overlays Opower’s business-as-usual HER with a rewards program that offers financial incentives for reductions in home energy use to nearly 200,000 households.<sup>3</sup>

In the field experiment, we randomly assign customers to one of three groups: (i) a true control group, (ii) a group that is only exposed to regular HERs and is ineligible to sign up for the rewards program, and (iii) a group that we encourage to participate voluntarily in the rewards program in addition to receiving HERs. Our design therefore allows us to identify whether the introduction of the rewards program affects the manner in which households subsequently respond to the HER. In addition to exploring complementarities, we believe it is important to examine this possibility given prior work showing that financial rewards can crowd out non-pecuniary motives by assigning a “price” to a previously unpriced behavior (see, e.g., Gneezy et al., 2011, Bowles and Polania-Reyes, 2012, and Kamenica, 2012, for overviews of this literature). Furthermore, the opt-in nature of the rewards programs allows us to describe the characteristics of customers who actively chose to participate. Lastly, a comparison of households exposed to the combined intervention to those only receiving the baseline HER affords conclusions about subsequent usage for program participants.

We first conduct a traditional evaluation of the HER trial using data from all households that were assigned to repeated receipt of the report, including those also encouraged to participate in the rewards program. Our findings are consistent with previous work and confirm the stylized facts in the literature. In our pooled sample, treated households reduce energy

---

<sup>3</sup> Under the rewards program, households earn points based on changes in monthly energy use. Points earned through this program can be redeemed to purchase goods via an online portal. The range of goods available includes gift cards to popular companies like Starbucks and Amazon, so-called Tango cards (a form of digital currency), and donations to charities like Habitat for Humanity. See Figure A4 in the Appendix for an example. The program is akin to peak time rebates and other energy rebate programs (Wolak, 2011; Ito, 2015).

demand by about 1.3 percent relative to the control group. Furthermore, we find that observed reductions are greater for households whose pre-experiment average daily use exceeded that of the median household and for those whose variance in month-to-month use exceeded that of the median household in our sample. Throughout the paper, we highlight the importance of these heterogeneous responses when interpreting results and take them as a benchmark for assessing the success of the combined intervention.

We next explore the extent to which the introduction of the rewards programs impacts the overall effectiveness of the HER program. To do so, we allow the effect of the HER to differ for those households in the treatment group that were offered the opportunity to enroll in the rewards program and those that were not afforded this opportunity. Results from this exercise provide the first evidence of potential complementarities amongst the rewards program and the baseline HER: the estimated reduction in average daily electricity use for households offered the rewards program is approximately 40% greater than that observed for counterparts who only received the monthly HER.

To better understand what drives these differences, we split our sample of treated households into two groups – (i) those who never enrolled in the rewards program and (ii) those who self-selected into the rewards program – and compare differences in daily energy use across these groups with counterparts from our control group.<sup>4</sup> Empirical results from this exercise further strengthen the case for complementarities amongst the rewards program and the baseline HER intervention. The estimated reductions in daily energy use for customers who ultimately participate in the rewards program are more than double the approximate 1.3 percent reduction observed amongst the full sample of treated households. Moreover, the change in daily energy use for households that chose not to enroll in the rewards program is approximately 30 percent greater than responses of customers only exposed to the HER.

While these differences are interesting in and of themselves, our data are sufficiently rich to investigate which types of households self-select into the rewards program. We find that disproportionately many low-usage and/or low-variance households sign up for the rewards program. Such heterogeneity is noteworthy given past work showing that such types are least

---

<sup>4</sup> By construction, the households in group (i) include all of those who were assigned to the HER-only treatment and those who were offered the opportunity to enroll in the rewards program but elected not to do so.

responsive to the HER. In this regard, the data suggest a potential channel for the observed complementarity between the two interventions – they influence different parts of the customer distribution.

To better isolate the impact of the rewards program, we next study subsequent usage patterns of rewards households compared to those only receiving HER letters. For this purpose, we estimate intent-to-treat (ITT) and local average treatment effects (LATE) using the random encouragement design as an instrument for selection into the program. Although noisy due to low rates of enrollment in the rewards program, results highlight three interesting findings. First, the introduction of the rewards program led households to reduce monthly energy use by more than that observed amongst counterparts that only received the HER. Specifically, our ITT estimates suggest that the *marginal* effect of the rewards program is about twenty percent of the size of the baseline HER effect. Second, LATE estimates suggest that participation in the rewards program leads to an additional five percent reduction in monthly use – a figure that is approximately four times greater than the estimated HER effect.

Third, we find evidence of heterogeneous responses to the rewards program across user groups. Both the ITT and LATE estimates for low variance users are *greater* than that observed for high variance counterparts. Similarly, we find that the difference in the effect of the rewards program across high and low user groups is less than the difference in the effect of the HER across these same user groups. Viewed in conjunction with the data on enrollment, these results suggest an important reason for the complementarity between the HER and the rewards program: financial rewards disproportionately attract and induce energy conservation from user groups whose behavior is least affected by social comparisons.

As a final piece of evidence, we evaluate the success of the program from two additional perspectives: (i) private cost-effectiveness and (ii) a partial welfare analysis. To do so, we rely on approaches in previous work and derive the cost to the utility of conserving one kWh (Allcott and Mullainathan, 2010). Depending on the underlying assumptions, we derive measures of cost-effectiveness between 1.82 and 1.95¢/kWh. These values compare favorably to a host of alternative energy-efficiency programs, standard HERs, and subsidy programs in other settings (Allcott and Mullainathan, 2010; Ito, 2015). Furthermore, we use estimates of marginal generation costs and marginal carbon emissions to conduct a partial welfare analysis (Graff Zivin

et al., 2014). We find that welfare is likely to increase for any reasonable range of marginal social costs. This is because the program is akin to an increase in the energy price of participants which narrows the gap between private and social marginal costs in the service area of our partner utility.

Our findings can be interpreted as speaking to several distinct literatures. For the literature on the use of social comparisons or related “nudges” to manage residential resource use, our results shed light on the puzzle of how to increase conservation efforts amongst lower user groups and those with less month-to-month variation in use. The introduction of a rewards program that provides financial incentives for conserving energy disproportionately attracts such user groups and leads to subsequent reductions in energy use that exceed those realized through the receipt of a social comparison. More broadly, our results highlight the promise of carefully combining behavioral and financial incentives to achieve ambitious policy goals. In a policy environment characterized by an increasing number of smaller interventions such as nudges, it is important to understand how different incentives interact with each other and how suites of policies perform compared to their individual building blocks in isolation.

The remainder of the paper is structured as follows. In Section II, we describe the setting, experimental design, and data to our disposal. Section III presents the main body of evidence based on various empirical specifications. We provide additional heterogeneity analysis in Section IV. Section V derives policy implications before we conclude in Section VI.

## II. EXPERIMENTAL DESIGN

### A. *Set-Up*

We partnered with Opower to design a new rewards program to encourage energy conservation and evaluate the program’s impact using a natural field experiment with a utility in the US Northeast (see Harrison and List, 2004, on the various field experiment types). The program offers interested customers the opportunity to receive financial rewards for reductions in usage relative to a pre-specified baseline level.<sup>5</sup> These rewards are not direct monetary rebates but

---

<sup>5</sup> Each customer faces an individual, undisclosed baseline. Baselines are calculated based on a customer's usage for the same month in the previous year, and normalized by weather (heating degree days and cooling degree days). The use of an undisclosed baselines reduces the possibility that subjects distort behavior in the pre-intervention period as

rather accumulate automatically as points if usage drops below baseline – a reduction of one kilowatt hour (kWh) is worth one rewards point. As such, the program shares similarities with peak price rebates and other subsidies for reductions in usage below a baseline level (Faruqui and Sergici, 2010; Wolak, 2010; 2011; Ito, 2015). These types of subsidies create asymmetric incentives because only usage below the baseline is subject to increased marginal prices while increases above the baseline are not penalized and remain priced at the original level. This asymmetry introduces an “option to quit” or “giving up effect” (Wolak, 2010; Borenstein, 2013). We further acknowledge that the program design does not provide all features of a first-best Pigouvian solution. Nevertheless, this type of program offers an attractive and widely-applied alternative for regulators and utilities who are concerned about the political environment and customer satisfaction (see, e.g., Wolak, 2010; Borenstein, 2013; Ito, 2015).

Points earned via the rewards program could be redeemed to purchase goods like gift cards in an online portal at an exchange rate of approximately one cent per point.<sup>6</sup> To put this value in perspective, customers in the experimental population faced a base flat rate of 6.963 ¢/kWh in the year of the intervention which translates into the reward being equivalent to an approximate 14.4 percent subsidy on energy conservation.

[ Insert Figure 1 About Here ]

The rewards program is offered in conjunction with Opower’s existing *Home Energy Reports* (HERs; see, for example, Allcott, 2011; 2015). HERs are printed letters consisting of three main modules: (i) social comparison of a households’ monthly electricity usage to the average usage of 100 similar households (the “neighbor group”) and to the 20<sup>th</sup> percentile of usage within the same group (the “efficient group”), (ii) graphical information about the household’s usage trend over time, and (iii) a tip sheet with a list of more or less costly ways to reduce energy use in the home. See Figure 1 for an example.

---

a way to influence subsequent rewards; an important lesson learned in early pilot experiments testing critical peak pricing plans (Wolak, 2010).

<sup>6</sup> Figure A4 in the Appendix presents a screenshot of the rewards portal. Examples of goods that can be purchased with rewards points include gift cards (Amazon, Starbucks, etc.), donations to charities (e.g. Habitat for Humanity), and Tango Cards, a form of digital rewards card that can be used at dozens of stores. The exchange rate is not an exact mapping because larger items are discounted in terms of point costs. For example, a \$5 Starbucks gift card costs 475 points.

[ Insert Figure 2 About Here ]

Unlike the standard opt-out design for HERs, Opower and the partner utility decided to employ an opt-in approach for the trial intervention to minimize customer complaints. We use this decision to our advantage and develop a random encouragement design that allows us a more nuanced understanding of the program. Figure 2 summarizes the design. We chose this approach with four goals in mind: (i) to derive a clean measure of the impact of HERs on use, (ii) to understand how these impacts are affected by the introduction of the rewards program, (iii) to provide evidence on the customer types attracted by financial incentives through self-selection into the program, and (iv) the ability to evaluate subsequent changes in energy usage due to program participation. To achieve these goals, we randomly assigned customers to one of three treatment arms:

***Control:*** a true control group that never receives any correspondence from Opower

***HER Only:*** a group of households that only receives monthly HERs but is ineligible to participate in the rewards program

***Rewards Incentives:*** a final group of customers that receives monthly HERs identical to HER Only but is also offered the opportunity to enroll in the rewards program

[ Insert Figure 3 About Here ]

Customers in the third group receive information about the rewards program and are offered an initial balance of 150 points (or \$1.50) should they enroll in the program.<sup>7</sup> Once enrolled, points accumulate automatically for savings each month. To highlight this financial element, the marketing module includes the sentence “*Earn points for every kWh you save and **get rewarded***” which is displayed next to examples of goods that can be purchased through the online portal. In addition, the module includes a link to the registration page and prominently highlights the signup bonus.<sup>8</sup> See Figure 3 for an example of the encouragement module.

---

<sup>7</sup> Average monthly points earned for reductions in usage are approximately 60. Hence, our signup bonus is equal to two and a half months of savings, on average.

<sup>8</sup> Some Rewards Incentives customers face marketing modules that utilize several behavioral framings for the same program and signup bonus. Although a very important question, we leave the analysis of this variation for future work and focus on the general impacts of the rewards program in this study.

[ Insert Figure 4 About Here ]

We implement the experiment identically across two new deployment waves of Opower’s monthly HER. The timeline is depicted in Figure 4. For each household in the experiment, we observe twelve months of pre-experiment usage reads starting in March 2012.<sup>9</sup> In March 2013, Opower delivers the first HER to all households except Control. After two months of receiving standard HERs, Rewards Incentives homes receive the encouragement module as part of their third HER in May 2013. This module features prominently on the front page of the HER. Customers in Control never receive any information about the rewards program and are not eligible to participate even if they learn about it through other channels.

The rewards module is *only included in the third letter*. Due to a relatively low participation rate in the month after receipt of the modified HERs, Opower decided to run three subsequent email campaigns promoting the program in June, July, and August 2013. These emails use the same content and identical incentives as the HER marketing module and are sent to all Rewards Incentives customers who did not sign up in the first 31 days. We observe the date of sign up for customers throughout all encouragement campaigns.

For the remainder of the paper, we label households signing up during the first month *HER participants* and those who participate after receiving emails *Email participants*. Importantly, this distinction allows us to test differences between early adopters and households attracted by later emails. All letters follow Opower’s standards and emails are sent by an official Opower email address in professional format and design, ensuring credibility of the intervention.

## ***B. Sample and Data***

We observe monthly electricity usage for all customers from March 2012 to May 2015. There are two forms of attrition in the sample. First, households can actively opt out by contacting a telephone hotline and asking to be removed. Only 1.05 percent of households in our sample do so. Second, some households move out of their homes at some point after the first HER is received. Overall, approximately 14 percent of households move during the sample period, or about 5 percent per year. For these homes, we observe monthly use only until move-

---

<sup>9</sup> This set-up is necessary to construct HER content and household-level baselines. Opower follows this principle in all of their trials.

out and we are unable to track households to their new location. Regression analysis shows that move-out is uncorrelated with treatment assignment and pre-treatment usage is perfectly balanced across groups. Consequently, we are not concerned about attrition and include homes that become inactive in the main specifications.<sup>10</sup>

The two deployment waves differ along observable characteristics and exist for logistical reasons. Wave 1 consists of dual-energy customers, i.e. customers who use both gas and electricity. Gas is traditionally used for space-heating, water heating, and cooking and thus reduces baseline demand for electricity. Wave 2 exclusively contains electricity-only customers with greater baseline use. Furthermore, wave 2 households have higher income and larger families, on average. Randomization is implemented on the wave level and the final assignments are presented in Figure 2. The randomization procedure balanced on pre-experiment usage and we find that both waves are perfectly balanced in terms of all observables with the exception of the number of children (see Table A1 and Table A2 in the Appendix).

Overall, the experimental sample entails about 196,000 customers, 79,000 of which are in wave 1 and 117,000 in wave 2 (see Figure 2; Figure A1 in the Appendix presents the geographical distribution of households in the experiment). Together, these subjects combine for close to seven million household-month observations of average daily energy usage. We pool both waves to increase power but control for different baseline uses by including wave fixed effects.<sup>11</sup>

### **III. EXPERIMENTAL RESULTS**

This section presents the main results for three questions afforded by the experimental design. First, we investigate how HERs affect customers' energy demand, relate our findings to stylized facts from previous work, and explore how responses to the HER differ across particular subgroups. Second, we study the extensive margin and document the types of households that select into participation in the rewards program. Third, we exploit our randomized

---

<sup>10</sup> Exclusion of movers and/or opt-out households does not affect qualitative results but reduces statistical precision due to a smaller sample size. These results are available upon request.

<sup>11</sup> We also perform analyses on the wave level to ensure robustness of results. Results are very similar and available upon request. Furthermore, we run regressions allowing for month-of-sample fixed effects to differ across both waves for all main specifications in the paper. These results are presented in the Appendix.

encouragement design to estimate the impact of financial rewards on subsequent patterns of use – the intensive margin – using both ITT and LATE approaches.

Before presenting our main findings, we provide a brief overview of the success of the random encouragement design. Overall, 7,634 customers or about five percent of the eligible sample voluntarily participated in the rewards program; 1,238 in response to the HER marketing modules and 6,396 after receiving encouragements through emails.<sup>12</sup> Compared to the group of households only being offered a signup bonus, exposure to additional behavioral framings increased take-up by up to 1.5 percentage points, an effect that is highly significant ( $p < 0.001$ ).

### A. Home Energy Reports

As a first step, we evaluate the HER campaign following an extensive body of work (e.g., Allcott, 2011; Costa and Kahn, 2013; Ferraro and Price, 2013; Allcott and Rogers, 2014; Allcott, 2015). This literature, which explores behavior across a variety of unique sites and experiments, highlights two stylized facts. First, despite some variation in point estimates, the receipt of social comparisons generates reductions in use that typically range from one to two percent relative to a control group. Second, households with high levels of baseline usage demonstrate stronger response to such programs whereas treatments effects for households from the left tail of the usage distribution are negligible. In summary, social comparisons induce moderate conservation efforts concentrated amongst a particular subset of consumers.

To derive the effect of HERs on average daily usage, we compare Control to all customers receiving HERs, i.e. HER Only and Rewards Incentive households. We do so by performing an ordinary least squares estimation in the spirit of Allcott and Rogers (2014), utilizing data from the treatment period only, i.e. after the first report was delivered:

$$(1) \quad Y_{imw} = \alpha + \delta^T H_i + \beta_1 Y_{im}^{Pre} + \mu_m + \omega_w + \varepsilon_{imw}$$

where  $Y_{imw}$  is electricity demand in average kWh per day by household  $i$  in month-of-sample  $m$  and wave  $w$ .  $H_i$  is a binary indicator for assignment to receipt of HERs at the household level.  $\delta^T$  is the coefficient of interest and describes the average treatment effect (ATE) of receiving

---

<sup>12</sup> This difference is not surprising because HER modules were only included in one month; emails were sent out three months in a row. Furthermore, the email campaign only utilized the most successful subset of behavioral framings to maximize participation.

HERs.  $Y_{im}^{Pre}$  is the average daily use in the pre-experiment period by household  $i$  in the same calendar month as month-of-sample  $m$ . We also include month-of-sample ( $\mu_m$ ) and wave ( $\omega_w$ ) fixed effects to control for shocks affecting usage common to particular months and to account for different baseline usage across the two waves. Heteroskedasticity-robust standard errors are clustered at the household level for all specifications. In alternative models, we interact treatment with a binary indicator for households above the wave-level median in terms of either average pre-experiment usage or variance of pre-experiment usage.<sup>13</sup>

[ Insert Table 1a About Here ]

Table 1a presents results from the main specification. Columns (1) and (2) utilize the full sample, columns (3) and (4) exclude households who participate in the rewards program at any point in time, and column (5) compares program participants to Control. As noted in column (1), we find that receipt of HERs decreases daily usage by about 0.32 kWh, on average (or 9.75 kWh/month at 30.5 days;  $p < 0.01$ ). In relative terms, this estimate implies a decrease in energy demand of about 1.3 percent compared to average Control usage in the treatment period. This aligns very well with the first stylized fact and previous findings in the literature (Allcott 2011, 2015). To place these reductions into perspective, effects are equivalent to treated households turning off three state-of-the-art CFL light bulbs for eight hours daily.

[ Insert Table 1b About Here ]

A look at the interacted models in Table 1b reveals that responses are mainly driven by households with high baseline usage and/or variance. Across both measures, households below the median (*Treatment* coefficient) reduce demand by only 0.13 to 0.17 kWh or about half of the overall ATE ( $p < 0.01$ ). High users, on the other hand, exhibit additional reductions of 0.27 to 0.36 kWh (coefficient on the interaction) – a marginal effect larger than the overall ATE.

---

<sup>13</sup> In the Appendix, we report results from a more traditional difference-in-differences approach and from a specification allowing month-of-sample fixed effects to vary by wave (Table A3a, Table A3b, and Table A4). Findings are virtually unchanged but the reported approach provides the most precision.

Together, these observations clearly are in line with the second stylized fact: reductions are predominantly driven by high users.<sup>14</sup>

We next explore the extent of how the introduction of the rewards program impacts the way in which households respond to the baseline HER intervention. To this end, we augment equation (1) and allow the HER effect to differ across HER Only households and those that also receive the opportunity to enroll in the rewards program. Results from this analysis provide the first evidence of a potential complementarity amongst these interventions. As noted in column (2) of Table 1a, reductions in average daily use for households that were offered the opportunity to enroll in the rewards program were approximately 0.10 kWh (or 40 percent) greater than those observed amongst counterparts that only received the monthly HERs.

Investigating differences across various subsamples, we find that exclusion of participants only reduces point estimates slightly. For example, as noted in Column (3) of Table 1a, the average treatment effect for the sample of households that did not participate in the rewards program corresponds to an approximate 0.297 kWh reduction in average daily use. This estimate is not statistically different from column (1) at conventional levels, indicating that observed reductions are not solely driven by participants. Moreover, as noted in column (4), reductions are actually greater for the subset of non-participants that were offered the opportunity to enroll in the rewards program but elected not to. Exploring the effect of the HER on participating households provides additional evidence of the program complementarity. Column (5) shows that the estimated treatment effect for such households is approximately 2.3 times greater than that observed for the sample of all households and approximately 2.5 times greater than that observed for the subset of non-participants.

### ***B. Characteristics of Participants***

A natural next step is to ask which types of customers select into the program and, if along observable dimensions, those participating in the rewards program differ from those who do not participate in the program. For this purpose, we compare characteristics across three groups: (i) eligible non-participants, (ii) HER participants, and (iii) Email participants. We use the same two usage measures as above—average pre-experiment usage and variance of pre-

---

<sup>14</sup> If we run the same model with finer usage bins, e.g. deciles, we see that effects increase weakly with decile. This is consonant with Allcott (2011) and Ferraro and Price (2013).

experiment use—and also investigate a range of demographics that could impact program participation.

[ Insert Figure 5a About Here ]

[ Insert Figure 5b About Here ]

Figure 5a provides a graphical overview of average pre-experiment usage for all three types. We further divide usage into overall average usage, the average in summer months (June-September), and the average in winter months (December-March). The left panel plots outcomes for wave 1 customers, the right panel for wave 2 customers only. Clearly, groups differ substantially in their pre-experiment usage behavior. Across all comparisons, HER participants, represented by light grey squares, are the lowest users. They are followed by Email participants (dark grey diamonds) which consistently show lower averages than non-participants (blue triangles). In wave 1, for example, the overall average usage of HER participants (17.6 kWh) is about 11.4 percent lower than non-participants' (19.87 kWh). Email participants lie in the middle (19.3 kWh) and use about 3 percent less than non-participants. Group differences are even more pronounced in wave 2 which features higher baseline usage due to its composition of electricity-only customers and therefore more margins for behavioral adjustments. Figure 5b draws the same conclusions for variance of pre-experiment use.<sup>15</sup>

We empirically test these differences by regressing average usage on indicator variables for HER and Email participants (see Table A5 in the Appendix).<sup>16</sup> For all comparisons, differences are significant at  $p < 0.01$ . In terms of other observables, we find that participants have higher income and score higher on a green affinity index provided by a marketing consultancy ( $p < 0.01$  for both comparisons). Point estimates also suggest that HER participants are more likely to be owners, have smaller families, and are more likely to invest in utility-

---

<sup>15</sup> Interestingly, the structure of the rewards program mechanically benefits high-variance households. Because increases in usage are not penalized but reductions accumulate rewards points, all else equal, higher variance leads to higher payoffs regardless of behavioral responses (Wolak, 2010; Ito, 2015). To mitigate some of the concerns, the partner utility capped monthly rewards at 300 points (or \$3). Furthermore, we focus on actual usage responses rather than earned points throughout the analysis. Nevertheless, it is surprising to observe such stark differences to this prediction.

<sup>16</sup> We include wave fixed effects and report heteroskedasticity-robust standard errors.

sponsored home improvements, although none of these differences are significant at conventional levels.

We next explore differences between households in the program and eligible non-participants. For this end, we assign each household to its wave-level decile in terms of both usage measures. We then plot the difference between the proportion of participants in a given usage decile and a uniform baseline in Figure 6. Consequently, if participation were independent of pre-experiment usage, we should observe a straight line at zero as 10 percent of participants should come from each decile. A positive difference—represented by bars above the uniform counterfactual—indicates a disproportionately large number of participants while bars below the zero-line mean that fewer than 10 percent of participants are drawn from a given decile. We show this comparison for HER participants (light grey) and Email participants (black outlines) separately.

[ Insert Figure 6 About Here ]

Inspection reveals striking patterns. In all cases, participants are not only drawn from below the median but rather from the lowest three deciles of pre-experiment usage. Conversely, the highest three deciles are underrepresented in the sample of participants. A comparison of HER to Email participants suggests that the former deviate much more from the uniform baseline. This suggests that it is the lowest user groups that elect to sign-up for the rewards program within the first 30 days of receiving the initial encouragement module.

Providing numbers, the first three deciles attract 9.4 (10.5) percent more HER participants for average usage (variance of usage) than predicted by the uniform baseline. These values are smaller for Email participants (2.4 and 3.1 percent) but paint the same general picture. On the other end of the spectrum, there are about 8.2 percent (7.2 percent) fewer HER participants than expected in deciles eight to ten and 2.5 percent (4.7 percent) of Email participants. To determine statistical differences across groups, we perform Chi-squared tests. The distributions are significantly different from uniform and from each other for both measures and all comparisons ( $p < 0.01$ ). A Kolmogorov-Smirnov test with the full distribution of pre-experiment usage and variance leads to the same conclusion.

Taken together, our data provide evidence that a disproportionately large number of low-usage and/or low-variance customers participate in the rewards program. Importantly, these are exactly the types of households that are least responsive to traditional HERs (see Table 1b; Allcott, 2011; Ferraro and Price, 2013).<sup>17</sup> Given the mounting evidence of differential effects of the HER across the usage distribution, this finding is highly policy-relevant. However, to conclude that the rewards program complements standard interventions in a meaningful way, we need to investigate whether participation actually leads to subsequent reductions in usage.

### *C. Subsequent Use of Participants*

In evaluating the impact of program participation on subsequent usage, we provide results from two approaches. First, we capture the behavioral response of a typical eligible household exposed to the encouragement campaign—irrespective of the actual participation decision—by estimating an intent-to-treat (ITT) effect. Given the voluntary nature of the program, this is the measure of program impact of chief interest to the implementing utility. Second, we note endogeneity concerns due to self-selection into the program and estimate a causal effect of participation on subsequent usage via an instrumental variables (IV) estimator common to this literature (e.g., Fowlie et al., 2015a,b). Specifically, we instrument for actual signup with random assignment to the encouragement campaigns and estimate a local average treatment effect (LATE) for *compliers*, i.e. households that voluntarily participate in the program.<sup>18,19</sup>

In the following analyses, we are interested in *marginal responses* net of the baseline effect of the HER. The experimental design provides a natural way to achieve this goal by restricting our sample to households in the HER Only and Rewards Incentives groups. By doing so, households in the HER Only treatment are the *de facto* control group to which those exposed to both interventions are compared.

---

<sup>17</sup> Differences in signup across HER and Email participants further suggest that a different type of marginal household is attracted by the two encouragement channels. Future research will provide a more in-depth treatment of this relationship.

<sup>18</sup> For a causal interpretation of  $\delta^{LATE}$  we need to invoke the exclusion restriction that households only change usage indirectly via participation, not directly due to reception of the RI letters (and emails). While we cannot empirically confirm this assumption, the short-lived and relatively weak nature of our intervention suggests that it is credible.

<sup>19</sup> In this context, there are no always-takers because only those households receiving rewards framings can actually sign up, i.e. we do not observe a single signup from HER-only customers. Put differently, we face the issue of one-sided non-compliance in the sense that not all treated units actually receive treatment (rewards points). No ineligible customers signed up for the program.

$$(2) \quad Y_{imw} = \alpha + \delta^{ITT} R_i + \beta_1 Y_{im}^{Pre} + \mu_m + \omega_w + \varepsilon_{imw}$$

Equation (2) presents the ITT model where  $R_i$  is a binary indicator for assignment to the rewards encouragements and all other variables are defined as in equation (1), i.e. we include controls for use in the same calendar month ( $Y_{im}^{Pre}$ ), month-of-sample fixed effects ( $\mu_m$ ), and wave fixed effects ( $\omega_w$ ). For the IV specification, instead of  $R_i$ , we use an indicator that equals one in the month of signup and in all following months and zero otherwise,  $SignUp_{im}$ . We instrument for participation with random assignment to an RI framing,  $R_i$ , and estimate a two-stage least squares model.  $\delta^{LATE}$ , the coefficient on  $SignUp_{im}$ , can be interpreted as the LATE described above. All specifications feature heteroskedasticity-robust standard errors which are clustered at the household level. We estimate these models for HER participants, Email participants, and all participants separately. For the HER group, we include observations in and after May 2013; for the Email group we begin one month later, i.e. when the first emails are delivered to customers in June.

[ Insert Table 2 About Here ]

Table 2 presents ITT and LATE estimates for all three groups. We observe negative point estimates across all groups, indicating reductions in energy demand compared to HER Only customers. In interpreting these results, we focus on the policy-relevant overall program impacts as presented in the last two columns. We find significant effects for both ITT and LATE at  $p < 0.1$ . Point estimates for the ITT are around one fifth (21 percent) of the average reductions induced by the HER. This is indicative of sizable *additional* reductions in energy demand for the average household.

The LATE shows that participants reduce consumption by 1.4 kWh or approximately 4.4 times the HER effect. Compared to the typical effect of the HER, this is a significant improvement in conservation efforts. Before proceeding, it should be noted that such reductions are even more impressive given that disproportionately many low-usage households comply with the encouragement treatments. This suggests that a proper counterfactual would feature lower average use than that observed for the control group as a whole. In that case, the percentage reductions attributable to the rewards program would be even greater.

We take these observations as further evidence of program complementarities. The financial rewards program engages a subset of customers whose behavior is largely unaffected by the HER. Furthermore, introduction of the rewards program does not appear to negatively affect the response of households that do not elect to participate in the program. However, due to small take-up rates, the introduction of the program does not appear to significantly move the needle in terms of overall reductions.

#### IV. HETEROGENEITIES

This section provides a closer look at the impact of the rewards program across different customer types. We have identified that disproportionately many low-usage and/or low-variance households select into the rewards program. However, demand reductions presented in Section III.C might solely be driven by the typical HER respondents, i.e. high-usage and/or high-variance customers. To shed more light on this open question, we construct subsamples based on pre-experiment usage behavior and assign households to either an above-median group (*High*) or a below-median group (*Low*) for the two usage measures.<sup>20</sup>

[ Insert Table 3 About Here ]

Table 3 presents results. We estimate equations (1) and (2) and the IV approach for High and Low users separately. In Panel A (B), we report outcomes based on average pre-experiment usage (variance of pre-experiment use). Several interesting patterns emerge. First, we confirm findings from Section III.A and show that High users respond substantially stronger to HERs than Low users (magnitude of 4.4 in Panel A; difference significant at  $p < 0.01$ ). Second, ITT and LATE reveal that the rewards program induces demand reductions from low-usage and/or low-variance households. Taking underpowered point estimates at face value, we find that High users subsequently reduce demand by more than Low users but the gap between the two household types narrows compared to the gap in the effect of the HER. Furthermore, reductions of participating Low households exceed the High users' response to the baseline HERs (0.62 to 0.52 kWh), indicating that the program causes policy-relevant conservation efforts.

---

<sup>20</sup> We perform similar analyses based on other observables (demographics). These models do not offer additional insights and we omit them for brevity. Results are available upon request.

Shifting our focus to Panel B, we find striking differences. High variance households respond much stronger to receipt of the HER, as expected. However, program participation has substantial and differential effects on low users. Our estimates suggest that the ITT for low variance customers is about 50 percent larger than the average ITT. Reductions of this magnitude are policy-relevant because the ITT is equal to almost one third of the average HER effect in Table 1a. Furthermore, the LATE provides similar insights: low variance compliers significantly reduce usage by almost 1.9 kWh, on average, a value that is about 35 percent larger than the overall LATE in Table 3. High variance compliers, on the other hand, only reduce their usage by approximately 0.87 kWh.

Unlike average usage, variance is a crude measure of the adjustments households already make prior to any intervention. For example, homes that strongly respond to exogenous factors like weather should exhibit higher variance, *ceteris paribus*. These customers, who likely are more aware of costless ways to mitigate energy demand, respond strongly to HER letters. However, interestingly, the rewards program realizes reductions from homes with lower variance. Financial incentives seem to induce conservation from low users that was not achieved by normative letters.<sup>21</sup>

Revisiting our initial results, we now can draw more nuanced conclusions. Our intervention not only attracts disproportionately many low-usage and/or low-variance households but we also observe substantial demand reductions from these participants. Evidence suggests that traditional HER letters and the rewards program in conjunction work better than either program separately. Households attracted by financial rewards incentives appear to be different types than those who respond strongly to normative messages, leading to complementarities of the two interventions.

## V. POLICY IMPLICATIONS

In this section, we aim to expand on our empirical findings by exploring the policy implications of the rewards program. We first utilize administrative data from the partner utility to construct a particular measure of program success: cost-effectiveness. This measure provides the paramount

---

<sup>21</sup> High usage households are more likely to be above the HER's usage comparison by construction. However, due to the nature of the neighborhood comparison groups, many low users also experience above-comparison usage. Consequently, while this might be part of the story it is unlikely to explain its full extent. We do not have access to the content of HERs and the comparisons individual households were exposed to over time.

decision criterion from the perspective of a budget-constrained utility having to comply with conservation goals. We then consider a partial welfare analysis in light of the incentive structure which, in essence, increases the marginal price of participants' usage below their benchmark (see Section II.A). Therefore, the fundamental question from a welfare perspective is how the marginal price faced by residential customers compares to the social cost of producing the marginal unit abated. We conclude by providing a broader interpretation of when policies such as the rewards program are likely to contribute to social welfare.

### *A. Cost-Effectiveness Calculations*

Cost-effectiveness is a widely-applied metric in policy evaluation (e.g., Allcott and Mullainathan, 2010; Ito, 2015). It represents the cost of conservation to the utility and is often expressed in  $\text{¢/kWh}$ . This criterion is generally applied by utilities to decide between several policy options to comply with conservation goals imposed by regulators. In the case of the rewards program, program costs consist of the financial signup bonus and repeated subsidy payments to households that reduce energy demand below their baseline.<sup>22</sup> Importantly, this measure only takes into account costs borne by the utility and ignores all other direct and indirect costs. Based on monthly administrative data provided by the partner utility, we can construct a total tally of points awarded to program participants. Furthermore, points have a constant exchange rate to the monetary value of redeemable products which allows us to express program costs in dollars.<sup>23</sup> On the other side of the equation, we use estimates from Section III.C to capture total conservation in kWh. Mirroring previous sections, we focus on additional conservation efforts net of reductions due to the receipt of HERs.

[ Insert Table 4 About Here ]

We derive cost-effectiveness for two scenarios: (S1) scaling up the intervention to the total experimental sample and (S2) evaluating the impact of actual participants. On the cost side, the average participant accumulates about 1,455 points by April 2015. This amounts to total

---

<sup>22</sup> Allcott and Mullainathan (2010) show that implementing a conventional HER program costs about \$7.48 per household-year. Correspondence with Opower shows that, outside of up-front programming expenses, providing the marketing modules in HERs and emails was costless to the utility. We do not have a measure of up-front costs for the implementation of the rewards program and ignore these fixed costs in the calculations.

<sup>23</sup> We assume throughout that 1 point is worth  $\text{¢}1$  despite discounts for costly items. Consequently, we underestimate program costs slightly if customers tend to choose more expensive items.

program costs of about \$111,100 or \$14.56 per participant (\$0.74 per eligible household).<sup>24</sup> On the conservation side, we use the ITT effect for S1 and the LATE for S2 combined with corresponding sample sizes. Total savings are then determined by multiplying the conservation coefficient ( $\hat{\beta}$ ) with the sample size ( $N$ ) and scaling the resulting total person-day savings by the average time in the program ( $T$ , 570 days). Outcomes of this exercise are reported in Table 4 and show savings of about 7.4 and 6.1 million kWh for the two scenarios, respectively. Similarly, we vary the cost measure,  $c$ , depending on the scenario. For S1, we use the average point cost per eligible households and for S2 the cost per actual participant. The last step is to divide total costs by total savings which leads to cost-effectiveness of 1.95 and 1.82 ¢/kWh in S1 and S2, respectively.

These results indicate that the rewards program is an attractive policy option compared to a host of other energy-efficiency programs (1.6-6.4¢) and even the standard HER (2.5¢) (Allcott and Mullainathan, 2010). Our measures are also similar to Ito (2015), who estimates cost-effectiveness of a general rebate program in California to be 2.5¢ in inland areas. Furthermore, when compared to the residential rate during the experimental period (6.96¢), we conclude that the program is a cost-effective strategy for the utility.<sup>25</sup>

### ***B. Welfare Considerations***

We next move from the perspective of the utility to that of a social planner by conducting a (partial) welfare analysis. In a first step to capture the welfare effects of energy-efficiency nudges, Allcott and Kessler (2015) use multiple price lists to elicit willingness-to-pay (WTP) of customers for continued receipt of HERs. In a revealed preference interpretation, such a measure includes otherwise unobservable indirect costs and benefits to customers (e.g., investments, time cost, psychological costs, warm glow). Allcott and Kessler (2015) find that, on average, WTP is positive and the HER increases social welfare. However, there is substantial heterogeneity across recipients and non-energy costs reduce welfare gains considerably. Nonetheless, the HER has attractive features from the point of view of the utility as well as the social planner.

---

<sup>24</sup> By the end of the sample period, only a small percentage of accumulated points was redeemed by participants (23 percent). This observation suggests that some customers might never actually turn virtual points into a real cost to the charity. Consequently, our back-of-the-envelope calculations might overstate actual program costs.

<sup>25</sup> We also obtain hourly wholesale market prices faced by the partner utility in its local load zone as an alternative measure of private costs to the utility for providing an additional kWh. The unweighted average price in 2013 was 5.61¢, the price weighted by load was 6.03¢. Conclusions are identical.

While our experiment does not provide the necessary variation to conduct an analysis akin to Allcott and Kessler (2015), we can utilize findings from previous work and knowledge of the underlying incentive structure to derive welfare implications. In particular, we ask the question of whether an increase in the marginal price faced by participants (P) is likely to increase or decrease welfare by comparing it to the marginal social cost (MSC) of electricity production. The structure of the rewards program implies that price changes are not experienced by all customers but rather by participants should they reduce consumption below some reference level. Yet, the program increases P for some customers and welfare implications depend on whether the original P was above or below MSC.

We construct MSC based on work in Graff Zivin et al. (2014), who estimate marginal generation costs and marginal carbon emissions for all NERC regions and hour-of-day.<sup>26</sup> Marginal costs vary substantially across regions and times within the US. The general intuition for this result is that timing and location of demand reductions can have very different effects depending on which generator's production is displaced on the margin (Holland and Mansur, 2008; Borenstein, 2012, Holland et al., 2016). Unfortunately, we do not have access to high-frequency data and cannot speak to the time dimension.<sup>27</sup>

Our measure of partial MSC combines unweighted average marginal generation costs for the NERC region of the partner utility (NPCC) from Graff Zivin et al. (2014; Table A3, p. 266) and marginal carbon emissions (Panel A of Fig. 5, p. 259) translated into dollar values by using current social cost of carbon estimates (\$40.45 per metric ton or 1.835 ¢/lb.).<sup>28</sup> Partial MSC for the region of our partner amounts to 8.27 ¢/kWh.<sup>29</sup> Importantly, this approach provides a lower bound on MSC as it does not include other pollutants such as sulfur oxide and particulate matter and other costs.

---

<sup>26</sup> Holland et al. 2016 take a very similar approach.

<sup>27</sup> If all reductions take place during peak demand hours, we likely underestimate welfare gains considerably while reductions primarily in off-peak imply that we overstate welfare gains.

<sup>28</sup> The social cost of carbon is extracted from the EPA (<https://www.epa.gov/climatechange/social-cost-carbon>) and we convert the 3 percent estimate from 2015 into 2013 dollars.

<sup>29</sup> Average unweighted marginal generation costs are 5.924 ¢/kWh and marginal carbon emissions are 2.349 ¢/kWh. These values are based on data from 2007-2009 used in Graff Zivin et al. (2014). We also obtain the wholesale market prices faced by the partner utility which provide very similar measures of private costs (unweighted average price in 2013 of 5.61 ¢/kWh; price weighted by load of 6.03 ¢/kWh) and lead to the same conclusions throughout.

To determine welfare impacts of the rewards program, we compare partial MSC to P with and without the subsidy for energy conservation. The flat rate at the beginning of the intervention in March 2013 was 6.96 ¢/kWh and the implied subsidy increases the *de facto* marginal price for program participants on units below the reference level to 7.96 ¢/kWh. From a welfare perspective, such an increase is beneficial if the MSC is above the private cost faced by customers. This is clearly the case for our partner utility.<sup>30</sup> Despite only considering partial MSC, the increase in P narrows the gap between private and social marginal costs without exceeding MSC.

Furthermore, following arguments in Boomhower and Davis (2014) and Ito (2015), utilities tend to pass through program costs to customers, implying a future increase in P for participants and non-participants alike. Our partial welfare analysis suggests that moderate rate hikes would lead to welfare increases in the case of our partner utility. More generally, welfare conclusions depend on the local cost structure – regions and times with MSC exceeding P imply increases, P greater than MSC implies decreases in welfare.<sup>31</sup>

## VI. DISCUSSION

Behavioral policies have become a workhorse for economists and policy makers in recent years. While such interventions have been shown to induce behavioral change at relatively low cost, they are not without limitations. Across several domains, including tax compliance, charitable giving, and reducing employee theft, social cues have been found to be important. For example, within the area of residential energy demand, social comparison letters have had import—with effect sizes of nearly 2 percent observed—but reductions are largely driven by households in the right tail of the usage (and variance) distribution across dozens of sites.

We use a natural field experiment to showcase a promising way to both increase treatment effect size and impact the entire consumer distribution. The core of our approach

---

<sup>30</sup> Two other utilities operate in the state of our partner. In 2013, the first utility charged 7.31 ¢/kWh, which implies welfare improvements if a similar rebate policy were being implemented. The second utility uses different rates depending on the season. From October to May, our calculations imply welfare gains from further price increases, for the June to September season P outweighs MSC and welfare would fall due to a larger gap between private and social cost.

<sup>31</sup> For instance, Ito (2015) shows that welfare conclusions depend on the tier a customer is in. Unlike California's tiered pricing schemes, customers in our experiment face a flat rate.

relies on complementarities between Opower’s traditional home energy reports and a novel program offering financial rewards for demand reductions. We find that complementarities arise through three channels. First, the rewards program attracts disproportionately many low-usage and/or low-variance participants. This is precisely the part of the customer distribution least responsive to Opower’s business-as-usual programs. Second, introduction of the rewards program does not negatively affect responses of non-participants, i.e. there is no crowd-out of conservation efforts. Third, estimates indicate sizable reductions after signup for all participating customer types. Hence, not only do the “correct” customers select into the program but they also reduce energy demand significantly. In our setting, a combination of the two interventions unequivocally increases environmental conservation compared to using either approach individually.

Despite these important complementarities, the combined intervention fails to move the needle significantly for the average household. The main reason for the modest average effect is low participation despite our offering of a financial sign-up bonus. While opt-in policies play an important role in policy making, economists still lack a clear understanding of how we can increase the success of voluntary programs (besides turning to defaults; e.g., Kahneman, 2003). We believe that there is much scope for future work harnessing insights from behavioral economics to increase participation rates. Nevertheless, the use of a random encouragement design affords us to provide insights otherwise unavailable – it acts as a screening device for customers interested in the program (e.g., Lazear et al., 2012).

More broadly, our natural field experiment provides a successful case study for combining popular behavioral and more traditional price-based programs to achieve ambitious policy goals. While multiple incentives have been shown to attenuate each other under some circumstances, the rewards program suggests the need for a better understanding of when incentives do and do not work well together (e.g., Gneezy et al., 2011). In a policy environment with an increasing number of small “nudges”, combining various interventions to carefully design a suite of policies can be a viable alternative to one-size-fits-all approaches. Future work should explore this question in greater detail.

## REFERENCES

- Allcott, Hunt. 2011. "Social Norms and Energy Conservation." *Journal of Public Economics*, 95(9): 1082-1095.
- Allcott, Hunt. 2015. "Site Selection Bias in Program Evaluation." *Quarterly Journal of Economics*, 130(3): 1117-1165.
- Allcott, Hunt, and Judd B. Kessler. 2015. "The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons" NBER Working Paper, No. 21671.
- Allcott, Hunt, and Sendhil Mullainathan. 2010. "Behavior and Energy Policy." *Science*, 327(5970): 1204-1205.
- Allcott, Hunt, and Todd Rogers. 2014. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." *American Economic Review*, 104(10): 3003-3037.
- Ayres, Ian, Sophie Raseman, and Alice Shih. 2013. "Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage." *Journal of Law, Economics, & Organization*, 29(5): 992-1022.
- Beshears, John, James J. Choi, David Laibson, Brigitte C. Madrian, and Katherine L. Milkman. 2015. "The Effect of Providing Peer Information on Retirement Savings Decisions." *Journal of Finance*, 70(3): 1161-1201.
- Boomhower, Judson, and Lucas W. Davis. 2014. "A Credible Approach for Measuring Inframarginal Participation in Energy Efficiency Programs." *Journal of Public Economics*, 113: 67-79.
- Borenstein, Severin. 2012. "The Private and Public Economics of Renewable Electricity Generation." *Journal of Economic Perspectives*, 26(1): 67-92.
- Borenstein, Severin. 2013. "Effective and Equitable Adoption of Opt-In Residential Dynamic Electricity Pricing." *Review of Industrial Organization*, 42(2): 127-160.
- Bowles, Samuel, and Sandra Polania-Reyes. 2012. "Economic Incentives and Social Preferences: Substitutes or Complements?" *Journal of Economic Literature*, 50(2): 368-425.

- Brent, Daniel A., Joseph H. Cook, and Skylar Olsen. 2015. "Social Comparisons, Household Water Use, and Participation in Utility Conservation Programs: Evidence from Three Randomized Trials." *Journal of the Association of Environmental and Resource Economists*, 2(4): 597-627.
- Costa, Dora L., and Matthew E. Kahn. 2013. "Energy Conservation “Nudges” and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment." *Journal of the European Economic Association*, 11(3): 680-702.
- Croson, Rachel, and Jen Shang. 2008. "The Impact of Downward Social Information on Contribution Decisions." *Experimental Economics*, 11(3): 221-233.
- Faruqui, Ahmad, and Sanem Sergici. 2010. "Household Response to Dynamic Pricing of Electricity: A Survey of 15 Experiments." *Journal of Regulatory Economics*, 38(2): 193-225.
- Ferraro, Paul J., and Michael K. Price. 2013. "Using Nonpecuniary Strategies to Influence Behavior: Evidence from a Large-Scale Field Experiment." *Review of Economics and Statistics*, 95(1): 64-73.
- Fowlie, Meredith, Michael Greenstone, and Catherin Wolfram. 2015a. "Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program." NBER Working Paper, No. 21331.
- Fowlie, Meredith, Michael Greenstone, and Catherine Wolfram. 2015b. "Are the Non-Monetary Costs of Energy Efficiency Investments Large? Understanding Low Take-up of a Free Energy Efficiency Program." *American Economic Review: Papers & Proceedings*, 105(5): 201-204.
- Frey, Bruno S., and Stephan Meier. 2004. "Social Comparisons and Pro-Social Behavior: Testing ‘Conditional Cooperation’ in a Field Experiment." *American Economic Review*, 94(5): 1717-1722.
- Gerber, Alan S., and Todd Rogers. 2009. "Descriptive Social Norms and Motivation to Vote: Everybody’s Voting and So Should You." *Journal of Politics*, 71(1): 178-191.

- Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel. 2011. "When and Why Incentives (Don't) Work to Modify Behavior." *Journal of Economic Perspectives*, 25(4): 191-210.
- Graff Zivin, Joshua S., Matthew J. Kotchen, and Erin T. Mansur. 2014. "Spatial and Temporal Heterogeneity of Marginal Emissions: Implications for Electric Cars and Other Electricity-Shifting Policies." *Journal of Economic Behavior & Organization*, 107(A): 248-268.
- Hallsworth, Michael, John A. List, Robert D. Metcalfe, and Ivo Vlaev. 2017. "The Behavioralist as Tax Collector: Using Natural Field Experiments to Enhance Tax Compliance." *Journal of Public Economics*, forthcoming.
- Harrison, Glenn W. & John A. List. 2004. "Field Experiments." *Journal of Economic Literature*, 42 (4), 1009-1055.
- Holland, Stephen P., and Erin T. Mansur. 2008. "Is Real-Time Pricing Green? The Environmental Impacts of Electricity Demand Variance." *Review of Economics and Statistics*, 90(3): 550-561.
- Holland, Stephen P., Erin T. Mansur, Nicholas Z. Muller, and Andrew J. Yates. 2016. "Are There Environmental Benefits from Driving Electric Vehicles? The Importance of Local Factors." *American Economic Review*, 106(12): 3700-3729.
- Ito, Koichiro. 2015. "Asymmetric Incentives in Subsidies: Evidence from a Large-Scale Electricity Rebate Program." *American Economic Journal: Economic Policy*, 7(3): 209-237.
- Kahneman, Daniel. 2003. "Maps of Bounded Rationality: Psychology for Behavioral Economics." *American Economic Review*, 93(5): 1449-1475.
- Kamenica, Emir. 2012. "Behavioral Economics and Psychology of Incentives." *Annual Review of Economics*, 4: 427-452.
- Lazear, Edward P., Ulrike Malmendier, and Roberto A. Weber. 2012. "Sorting in Experiments with Application to Social Preferences." *American Economic Journal: Applied Economics*, 4(1): 136-163.

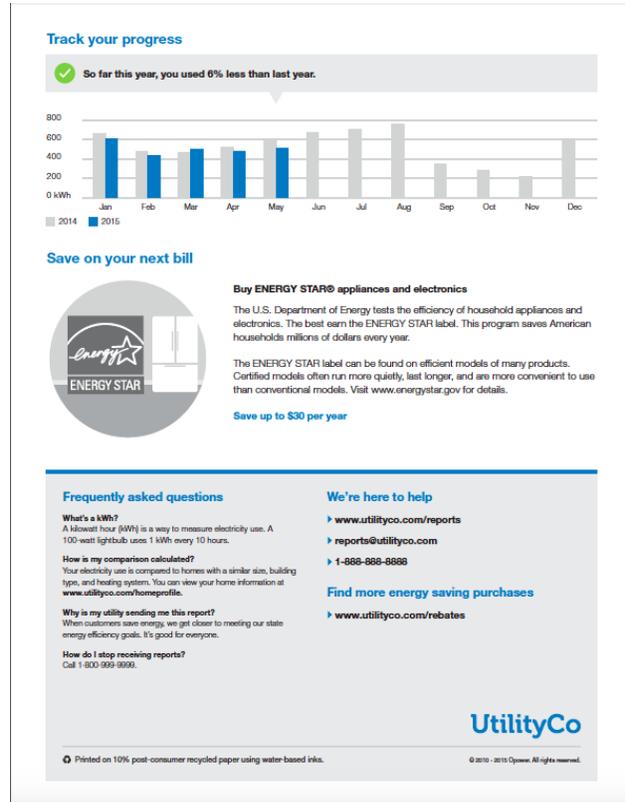
- Shang, Jen, and Rachel Croson. 2009. "A Field Experiment in Charitable Contribution: The Impact of Social Information on the Voluntary Provision of Public Goods." *Economic Journal*, 119(540): 1422-1439.
- Wolak, Frank. A. 2010. "An Experimental Comparison of Critical Peak and Hourly Pricing: The PowerCentsDC Program." Department of Economics Stanford University, Working Paper.
- Wolak, Frank A. 2011. "Do Residential Customers Respond to Hourly Prices? Evidence from a Dynamic Pricing Experiment." *American Economic Review: Papers & Proceedings*, 101(3): 83-87.

# Figures and Tables

Figure 1: Opower's Home Energy Report



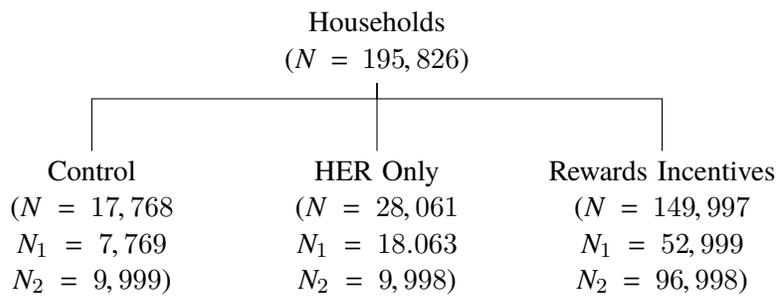
(a) Front



(b) Back

Notes: The two panels present a typical Home Energy Report generated by Opower. The front page provides the neighbor comparison and injunctive norm; the back page includes a personal usage comparison over time and conservation tips. Our marketing module was included in the lower half of the front page in May 2013. Source: Opower.

Figure 2: Experimental Design



*Notes:* Households are randomly assigned to one of three treatments within two deployment waves. Control customers do not receive any correspondence from Opower. HER Only customers receive monthly HERs beginning in March 2013. Rewards Incentives customers are encouraged to participate in the rewards program in addition to receiving monthly HERs.  $N$  depicts the overall sample size,  $N_1$  the number of customers per treatment cell in wave 1, and  $N_2$  the treatment assignment in wave 2. For evidence of a successful randomization, please consult Table A1 and Table A2 in the appendix.

Figure 3: Example Encouragement Message

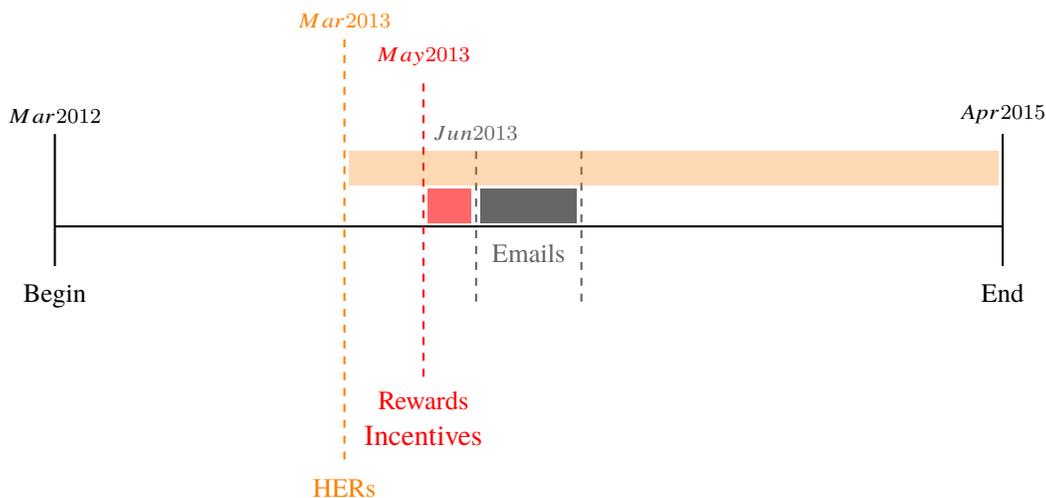
Earn points for every  
kWh you save and  
**get rewarded**



Earn 200 bonus points just for signing up: Visit [nationalgridRI/rewards](http://nationalgridRI/rewards)

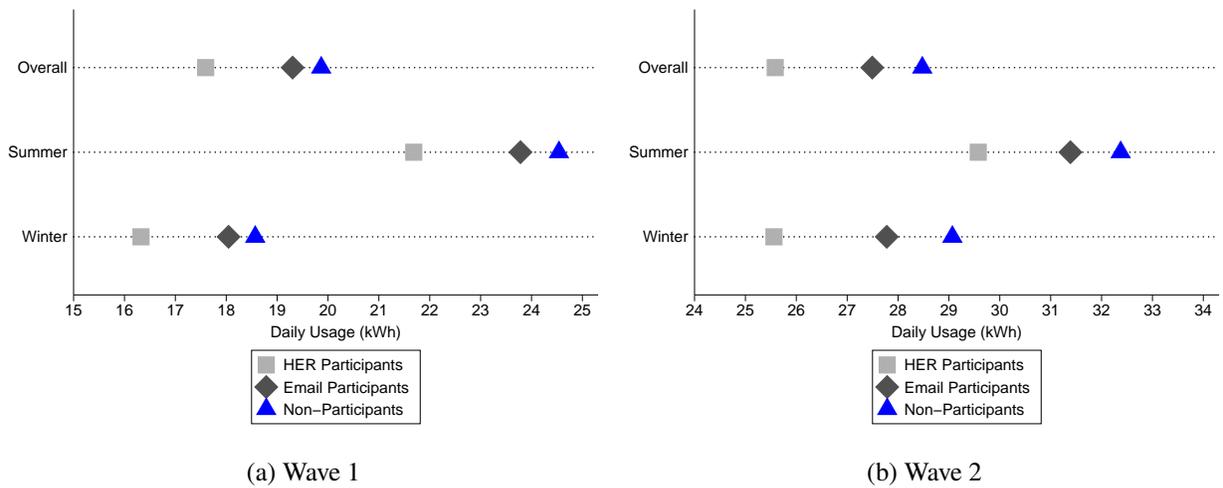
*Notes:* Content of an example encouragement module included in the third HER (May 2013) for customers in the Rewards Incentives treatment. The same content was used for encouragement emails in June, July, and August 2013 for Rewards Incentives customer who did not sign up in the first 31 days.

Figure 4: Timeline of the Experiment



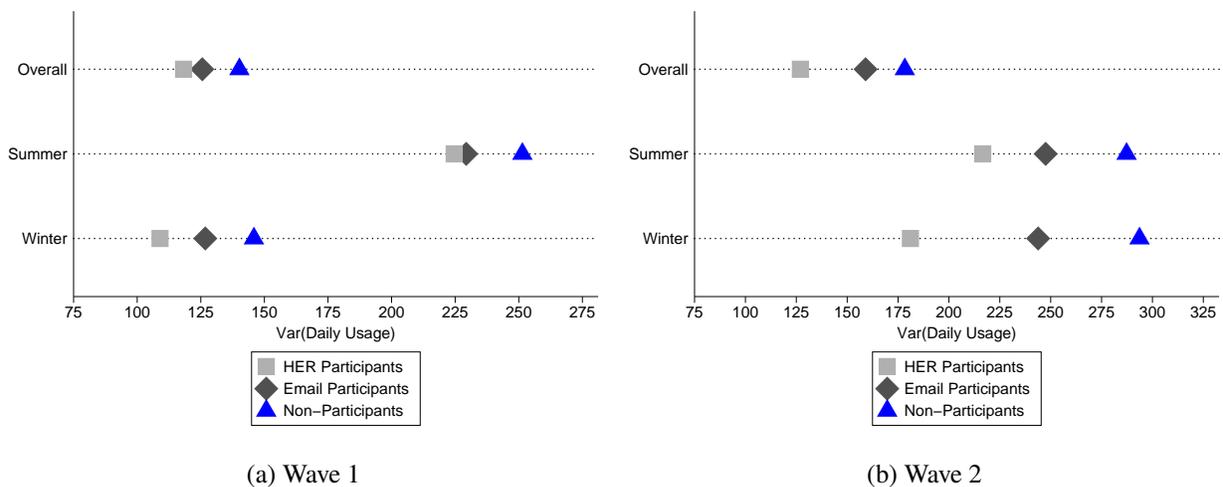
*Notes:* Vertical lines represent the begin dates of important interventions and rectangles of the same color represent the duration. We observe one year of energy usage before the first HER is delivered in March 2013. The marketing module for the rewards program was included in the May 2013 HER and consequent email campaigns were implemented in June, July, and August 2013. We observe average daily usage for each month until April 2015 for all customers in the experiment. This timeline is identical for both deployment waves.

Figure 5a: Differences in Average Usage between Customer Groups



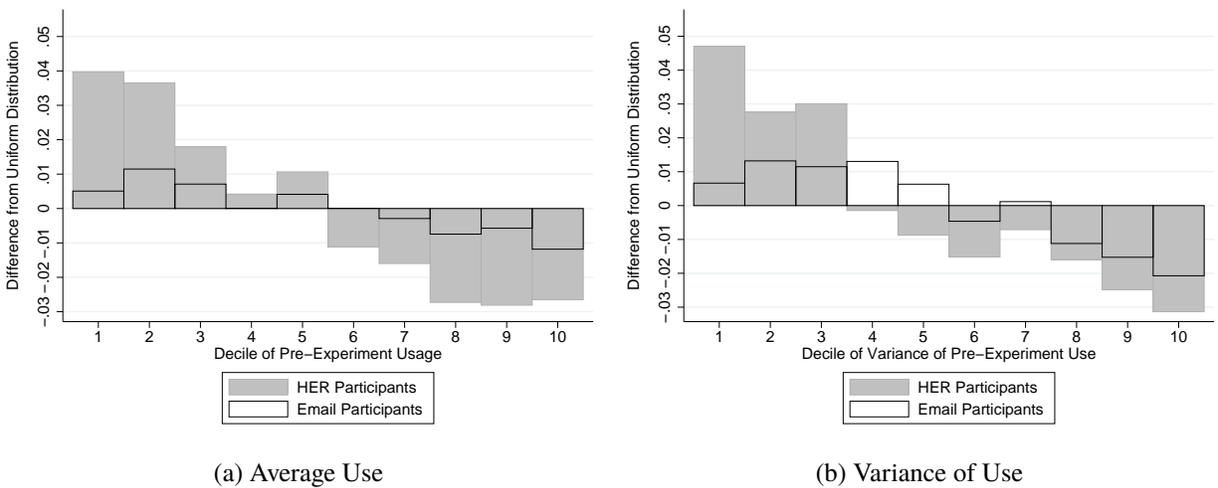
Notes: Average daily pre-experiment usage in kWh by deployment wave for three groups: i) HER participants, ii) Email participants, and iii) non-participants. Average usage is obtained separately for the entire pre-experiment period (March 2012-March 2013), summer (June-September), and winter (December-March) months. All differences are significant at a  $p < 0.01$  in a linear regression.

Figure 5b: Differences in Variance of Use between Customer Groups



Notes: Pre-experiment variance of daily usage in kWh by deployment wave for three groups: i) HER participants, ii) Email participants, and iii) non-participants. Average usage is obtained separately for the entire pre-experiment period (March 2012-March 2013), summer (June-September), and winter (December-March) months. All differences are significant at a  $p < 0.01$  in a linear regression.

Figure 6: Heterogeneity in Use: Deciles of Pre-Experiment Usage and Variance of Use



*Notes:* Difference between a uniform distribution and the actual proportions of participants in each decile of two usage behaviors: (a) average pre-experiment usage and (b) variance of pre-experiment use. We plot results by timing of signup. HER participants signed up during the initial HER campaign in May 2013, Email participants during subsequent email campaigns in June, July, and August 2013. The reference level is the uniform distribution across deciles, i.e. 10% of observations in each decile. Chi-Squared tests reject equal distributions for all comparisons at at  $p < 0.01$ .

Table 1a: Impact of Home Energy Reports on Use

	All Households		Non-Participants		Participants
	(1)	(2)	(3)	(4)	(5)
Treatment	-0.3158*** (0.0477)	-0.2311** (0.0989)	-0.2968*** (0.0478)	-0.2314** (0.0988)	-0.7350*** (0.0733)
Treatment · Rewards		-0.1011 (0.1124)		-0.0783 (0.1125)	
$R^2$	0.721	0.721	0.721	0.721	0.723
N	4,616,989	4,616,989	4,428,616	4,428,616	607,169

*Notes:* Dependent variable is average daily electricity usage (kWh) in a given month. All models include month-of-sample and wave fixed effects. In addition, we control for pre-experiment usage by including average daily use in the same calendar month before treatment. Heteroskedasticity-robust standard errors are clustered at the household level for all specifications. “Rewards” is a binary indicator equal to one for Rewards Incentives households. Columns (1)-(2) utilize the full sample, columns (3)-(4) exclude participating households, and column (5) restricts the sample to participants. We only present coefficients of interest and omit baseline differences and usage controls. Please consult Equation (1) and the following paragraph for details. \*\*\* denotes significance at the 1 percent level, \*\* at the 5 percent level, and \* at the 10 percent level.

Table 1b: Heterogeneous Impacts of Home Energy Reports on Use

	All Households		Non-Participants		Participants	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.134*** (0.045)	-0.169*** (0.045)	-0.126*** (0.046)	-0.160*** (0.045)	-0.295*** (0.065)	-0.357*** (0.064)
Treatment · High Usage	-0.365*** (0.095)		-0.345*** (0.095)		-0.867*** (0.150)	
Treatment · High Variance		-0.285*** (0.095)		-0.269*** (0.095)		-0.728*** (0.153)
High Usage	1.704*** (0.095)		1.690*** (0.096)		1.850*** (0.133)	
High Variance		1.196*** (0.091)		1.187*** (0.091)		1.254*** (0.098)
$R^2$	0.722	0.722	0.722	0.721	0.724	0.724
N	4,616,989	4,616,989	4,428,616	4,428,616	607,169	607,169

*Notes:* Dependent variable is average daily electricity usage (kWh) in a given month. All models include month-of-sample and wave fixed effects. In addition, we control for pre-experiment use by including average daily use in the same calendar month before treatment. Heteroskedasticity-robust standard errors are clustered at the household level for all specifications. “High Usage” describes a binary indicator for above-median average usage in the pre-treatment period (March 2012-February 2013), “High Variance” an indicator for above-median variance of pre-treatment usage. Columns (1)-(2) utilize the full sample, columns (3)-(4) exclude participating households, and columns (5)-(6) restrict the sample to participants. We only present coefficients of interest and omit baseline differences and usage controls. Please consult Equation (1) and the following paragraph for details. \*\*\* denotes significance at the 1 percent level, \*\* at the 5 percent level, and \* at the 10 percent level.

Table 2: Impact of Program Participation on Subsequent Use

	HER Participants		Email Participants		All Participants	
	ITT	LATE	ITT	LATE	ITT	LATE
Rewards	-0.0495 (0.0400)		-0.0640 (0.0409)		-0.0665* (0.0398)	
Sign-Up		-5.4340 (4.3975)		-1.5986 (1.0214)		-1.4027* (0.8394)
$R^2$	0.721	0.720	0.720	0.720	0.721	0.721
N	3,705,259	3,705,259	3,650,230	3,650,230	3,850,288	3,850,288

*Notes:* Dependent variable is average daily electricity usage (kWh) in a given month. All models include month-of-sample and wave fixed effects. In addition, we control for pre-experiment use by including average daily use in the same calendar month before treatment. Heteroskedasticity-robust standard errors are clustered at the household level for all specifications. Control households are excluded from the analysis. We present Intent-to-Treat (ITT) effects of being exposed to the encouragement campaigns (“Rewards”). Furthermore, we provide a Local Average Treatment Effect (LATE) based on an instrumental variables approach in which we instrument for actual participation with receipt of encouragements. Columns (1)-(2) present findings for HER participants, columns (3)-(4) for Email participants, and columns (5)-(6) for all participants. Please consult Equation (2) and the following paragraph for details. \*\*\* denotes significance at the 1 percent level, \*\* at the 5 percent level, and \* at the 10 percent level.

Table 3: Heterogeneous Impacts of Program Participation on Subsequent Use

	HER		ITT		LATE	
	High	Low	High	Low	High	Low
<i>Panel A: Average Pre-Experiment Use</i>						
Treatment	-0.5178*** (0.0826)	-0.1184*** (0.0459)				
Rewards			-0.0905 (0.0685)	-0.0319 (0.0382)		
Sign-Up					-2.0802 (1.5740)	-0.6188 (0.7421)
$R^2$	0.660	0.509	0.662	0.507	0.661	0.507
N	2,356,535	2,260,454	1,968,621	1,881,667	1,968,621	1,881,667
<i>Panel B: Variance of Pre-Experiment Use</i>						
Treatment	-0.4719*** (0.0835)	-0.1538*** (0.0448)				
Rewards			-0.0370 (0.0688)	-0.0990** (0.0396)		
Sign-Up					-0.8715 (1.6220)	-1.8882** (0.7564)
$R^2$	0.691	0.655	0.693	0.643	0.693	0.642
N	2,313,927	2,303,062	1,928,911	1,921,377	1,928,911	1,921,377

*Notes:* Dependent variable is average daily electricity usage (kWh) in a given month. All models include month-of-sample and wave fixed effects. In addition, we control for pre-experiment use by including average daily use in the same calendar month before treatment. Heteroskedasticity-robust standard errors are clustered at the household level for all specifications. Control households are excluded from the analysis. We present Intent-to-Treat (ITT) effects of being exposed to the encouragement campaigns (“Rewards”). Furthermore, we provide a Local Average Treatment Effect (LATE) based on an instrumental variables approach in which we instrument for actual participation with receipt of encouragements. Results are based on all participants. Households are assigned to the binary category “High” in Panel A (B) if their average pre-experiment usage (variance of pre-experiment use) is above the median within their wave and “Low” if it is below. \*\*\* denotes significance at the 1 percent level, \*\* at the 5 percent level, and \* at the 10 percent level.

Table 4: Cost-Effectiveness Calculations

	Scenarios	
	S1	S2
<i>Parameters:</i>		
$\hat{\beta}$ (kWh)	0.0665	1.4027
$N$ (Customers)	195,826	7,634
$c$ (\$)	0.741	14.559
$T$ (Days)	570	570
<i>Program Impacts:</i>		
Costs (\$)	145,107	111,145
Savings (kWh)	7,422,785	6,103,681
<i>Cost-Effectiveness:</i>		
$\phi$ /kWh	1.95	1.82

*Notes:* S1: use estimated ITT and average program costs per eligible household ( $c$ ) for all customers in the experiment; S2: use estimated LATE and average program costs per participant for all participants. To calculate costs, we use the observed average cost in \$ per household based on a conversion rate of 1¢/point. Total savings are calculated by multiplying the number of households ( $N$ ) by the corresponding average daily treatment effect ( $\hat{\beta}$ ) and the average number of days in the program for participants ( $T$ ). Lastly, cost-effectiveness is derived by dividing total costs and total savings. This measure can be interpreted as the cost to the utility (in ¢) of a reduction in demand of one kWh.