Social Norms and Energy Conservation

Hunt Allcott
MIT and NYU

August 24, 2009

Abstract

This paper evaluates a pilot program run by a company called Positive Energy to mail home energy reports that compare a household's energy use to that of its neighbors and provide energy conservation information. Using data from randomized natural field experiment at 80,000 treatment and control households in Minnesota, I estimate that the program reduces energy consumption by 1.9 percent relative to baseline. In a treatment group receiving reports each quarter, the effects appear to decay in the intervening months, suggesting that the reports successfully motivate or remind households to conserve but that this attention or motivation is not durable. I show that "profiling," or using a statistical decision rule to target the program at households whose observable characteristics suggest larger treatment effects, could substantially improve cost effectiveness in future programs. The effects of this program provide additional evidence that non-price "nudges" can substantially affect consumer behavior.

JEL Codes: C93, D12, L51, L94, Q40, Z13.

Keywords: Social norms, household energy conservation, randomized field experiments.

I thank, without implicating, Ian Ayres, Bob Cialdini, Tyler Curtis, Rajeev Dehejia, Kenny Gillingham, Larry Goulder, Matt Harding, Kosuke Imai, Seema Jayachandran, Karthik Kalyanaraman, Ogi Kavazovic, Alex Laskey, Aprajit Mahajan, Sendhil Mullainathan, Dave Rapson, Todd Rogers, Eldar Shafir, and seminar participants at Stanford University for helpful conversations related to this project.

1 Introduction

Climate change has emerged as one of the most pressing issues of the early 21st century, and energy efficiency could be a principal way of addressing it. Many analysts believe, however, that over the past 35 years energy efficiency has not lived up to its potential to deliver large energy savings at seemingly low cost. One critical opportunity that may be underexploited is insights from behavioral science, including the power of social norms. One critical difficulty has been a lack of consensus on how to measure the causal effects of energy efficiency programs, which has hindered the ability to learn what works and take demonstrably-effective programs to scale.

This paper econometrically evaluates a large-scale energy efficiency program run by a company called Positive Energy for an electric utility in Minnesota. The program involves sending to residential consumers Home Energy Reports with two key features. The first is an Action Steps Module that provides information, specifically targeted to each household, on ways to conserve energy. The second is a Social Comparison Module that details the household's electricity consumption and compares it to that of its one hundred nearest geographical neighbors in houses of comparable square footage. The normative comparison feature was motivated by academic work showing that social comparisons can powerfully affect behavior, including inducing people to conserve energy (Goldstein, Cialdini, and Griskevicius 2008), vote (Gerber and Rogers 2009), or stop littering (Cialdini, Reno, and Kallgren 1990). This literature has developed alongside work by economists on social learning[1] and conditional cooperation in the private provision of public goods[2].

Positive Energy's Home Energy Reports are but one example of a wide variety of energy efficiency programs that utilities operate to satisfy regulatory requirements, including information campaigns, energy audits and weatherization, and rebates for purchasing energy efficient durable goods such as lightbulbs, air conditioners, and water heaters. The causal effects of these programs are typically estimated using the "deemed savings approach": multiply the number of participants by ex-ante "engineering estimates" of the energy savings per participant relative to some counterfactual. While those in industry have long questioned the accuracy and precision of deemed savings and other approaches[3] and are aware of the conceptual usefulness of randomized controlled trials[4], there is no consensus on how the effects of energy efficiency programs should be measured. In this context, one of the remarkable features of the Positive Energy program is that it was implemented as a randomized, controlled experiment, allowing an unbiased estimate of treatment effects in the eligible population.

The effects of Positive Energy's program are of interest for two reasons. First, independent estimates of the reductions driven by the Home Energy Reports are important *per se*. The program has been introduced at utilities serving 15 percent of the U.S. population, in Northern and South-

[1] This literature includes Banerjee (1992), Beshears, *et al*, (2009), Conley and Udry (Forthcoming), Duflo and Saez (2002, 2003), Foster and Rosenzweig (1995), and Mobius, Niehaus, and Rosenblat (2005). In these settings, agents make a choice under uncertainty and draw inference from others' behavior because others may have distinct and useful information.

[2] Recent work on conditional cooperation includes Fischbacher, Gachter, and Fehr (2001, Shang and Croson (2004), Frey and Meier (2004), and Alpizar, Carlsson, and Johansson-Stenman (2008). These studies suggest that people are more likely to contribute to public goods when informed that others are contributing more.

[3] Nadel and Keating (1991) show that there was substantial bias and variance in deemed savings estimates relative to ex-post econometric evaluations.

[4] Energy efficiency programs are rarely evaluated using randomized trials, although the program analyzed by Davis (2008) is one exception. There is, however, a long history of randomized evaluations of energy *pricing* programs, including Aigner (1984), Allcott (2009), Aubin, *et al*, (1995) and Wolak (2006). Indeed, the British Domestic Tariffs Experiment, carried out between 1966 and 1972, was one of the first large-scale social experiments in any domain (Levitt and List 2009).

ern California, Washington, Minnesota, Illinois, Colorado, and Virginia, and it has been covered in the New York Times (Kaufman 2009), the Atlantic (Tsui 2009), National Public Radio, and other popular media outlets. Other utilities are considering adopting the program, and credible documentation of the magnitude of it effects will affect the disposition of millions of dollars in potential investment.

The success or failure of the Positive Energy program is also of more conceptual interest. Whether this program is perceived as successful will influence whether future energy efficiency programs are influenced by findings from behavioral science and evaluated via randomized trials. If the program works, it would be a remarkable illustration of the effectiveness of non-price interventions, in a world where the bulk of spending on energy efficiency programs currently goes to large rebates for household durables (Gillingham, Newell, and Palmer 2006). A successful non-price intervention it would be consistent with a growing empirical literature on how non-price interventions, or "nudges," can substantially affect individuals' behavior, including Bertrand, *et al*, (2010)[5], Benartzi and Thaler (2004), Ashraf, Karlan, and Yin (2006), and others.

The point estimates of the Average Treatment Effect (ATE) of Positive Energy's monthly Home Energy Reports range from a 1.8 to a 2.1 percent reduction in electricity consumption relative to baseline. The 95 percent confidence intervals in the five primary specifications span a range from 1.52 to 2.50 percent. Because of the large sample size and randomized control group, the estimated ATEs are highly robust to alternative specifications of fixed effects and control variables. The effects do vary somewhat depending on the envelopes in which they are delivered, the time elapsed since the beginning of the treatment, and the season.

Program households were randomly assigned to groups that would receive Reports every month versus every quarter. Interestingly, the Quarterly group treatment effects decay in subsequent months relative to the Monthly group, then increase again with the receipt of a second report, then decay again. The decay is economically significant, as it constitutes half of the treatment effect. Additional survey evidence suggests that this is the result of a cycle in which receiving a letter reminds or motivates the household to conserve, the household tires of the change over time, and motivation or memory is re-instilled upon receiving the next quarter's Report.

Econometric results also show that households that were high energy consumers before the treatment conserve substantially more than households whose baseline consumption was low. As I will discuss, this is consistent with one of the chief concerns in providing information on social norms: while individuals worse than the norm may improve, a "boomerang effect" might cause those better than the norm to regress (Clee and Wicklund 1980, Ringold 2002). More generally, I show that the program has heterogeneous treatment effects, not just as a function of baseline usage, but also correlated with other household characteristics observable to a program designer. This suggests that "profiling," or targeting future treatment toward units with higher expected treatment effects, could improve average efficacy and cost-effectiveness for those treated.

The final section of the paper builds on this insight to develop a statistical treatment rule under

[5] For their study, Bertrand, *et al*, partnered with one of the largest banks in South Africa to offer new loans to existing clients, via letters that varied both the interest rate offer and other psychological cues. They varied the number of different potential loans that were presented (to test whether greater choice could overload decision-making), how the interest rate was compared to some market benchmark, the race and gender of the person in a photo on the offer letter, the expiration date of the offer, whether the offer is combined with a promotional giveaway, and whether the letter mentions suggested uses for the loan. Consistent with economic theory, they found that consumers that had been offered lower interest rates were much more likely to take up the loans. They also found, however, that any one psychological cue could affect takeup by almost as much as a one to two percentage point change in the monthly interest rate. See Allcott and Mullainathan (2009a and 2009b) for further discussion of applications of non-price approaches to energy conservation suggested by recent research in psychology and economics.

which a decisionmaker allocates future treatment to minimize the expected cost of achieving some required aggregate treatment effect, conditional on information from the existing experiment. The routine draws on a recently-growing literature on using the results of randomized experiments to develop statistical decision rules for future treatments, including Berger, Black, and Smith (2000), Dehejia (2005), Graham, Imbens, and Ridder (2009), Hirano and Porter (2006), Imai and Strauss (2009), and Manski (2004, 2009). I show that, if the program were to be administered to only half of the population, profiling would improve cost effectiveness by 80 percent relative to random assignment.

This paper proceeds by providing background on Positive Energy's pilot experiment in Minnesota. Section III details the econometric strategy, Section IV presents results, Section V details the statistical treatment rule and gains from profiling, and Section VI concludes.

2 Experiment Overview

2.1 Motivating Literature

Social scientists have long been aware of the potential for social norms to affect individual behavior. For example, there is a large experimental literature in psychology that shows that providing people with information on social norms can have powerful influence (Cialdini 2003, Goldstein, Cialdini, and Griskevicius 2008, Gerber and Rogers 2009, Cialdini, Reno, and Kallgren 1990). This could occur for a number of reasons. We might emulate the norms of high status people to signal high status (Veblen 1899, Pesendorfer 1995). We might conform to customs due to social penalties for noncompliance (Akerlof 1980). We might follow a subgroup norm to signal horizontal type (Levy 1959, Wernerfelt 1990) or because of intrinsic utility from conforming to an identity (Akerlof and Kranton 2000). We might conform to a norm of prosocial behavior to signal benevolent underlying preferences (Bernheim 1994) or to otherwise receive social acclaim (Becker 1974). Finally, we might follow others because their choices are informative when we have imperfect information (Banerjee 1992). The explanations for conformity depend most importantly on whether the action taken is observed or unobserved and whether we are considering consumption of private goods or private provision of public goods.

There are three pathways through which information on neighbors' electricity consumption would most likely affect a household's quantity of electricity demanded. First, the Reports could induce households to conserve if individuals derive utility from being shown to be more virtuous than their neighbors. Second, because some externalities, primarily from power plant greenhouse gas emissions, are not internalized in electricity prices, many consumers perceive that energy conservation helps provide a public good (more moderate global climate). A literature on "conditional cooperation," including Alpizar, Carlsson, and Johansson-Stenman (2008), Cialdini (2003), Fischbacher, Gachter, and Fehr (2001), Frey and Meier (2004), and Shang and Croson (2004), has shown that people are more likely to contribute to public goods when informed that others are contributing.

Third, the information in the Home Energy Reports may facilitate social learning, as in Conley and Udry (Forthcoming), Duflo and Saez (2002, 2003), Foster and Rosenzweig (1995), Mobius, Niehaus, and Rosenblat (2005), and other applications. Electricity costs the average household about $1000 per year, but households may not be very knowledgeable about the amount of electricity they consume or what factors influence consumption. Given this uncertainty, new information on neighbors' consumption might induce the household to re-optimize around the level of household energy services or the efficiency with which energy input is transformed into energy services. For example, a household that learns that comparable households are using much less energy might infer that they themselves have low-cost opportunities to conserve.

All of these channels suggest that revealing the social norm should affect high consumption households more than low consumption households. Importantly, the second and third channels suggest that learning that they consume less than normal would induce low consumption households to conserve *less*. This has been called the "boomerang effect" (Clee and Wicklund 1980, Ringold 2002). In a pilot study of electricity conservation that influenced the design of Positive Energy's Home Energy Reports, Schultz, Nolan, Cialdini, Goldstein, and Griskevicius (2007) found that revealing the social norm did indeed cause low consumption households to increase consumption. Their solution was to add "injunctive social norms," which label others' behavior as good or bad, along with the "descriptive social norms," which simply describe others' behavior.

There is also a large body of existing work on how information provision and behavioral interventions can induce households to conserve energy. Giving consumers feedback on their consumption, providing information on energy savings opportunities, comparing their use to their neighbors' use, facilitating public or private goal setting, and structuring commitment devices have caused households to reduce energy consumption by 5-20 percent. Abrahamse, et al, (2005), Darby (2006), Fischer (2008), Shippee (1980), and Stern (1992) provide reviews of this literature. Many of these approaches, however, have been tested only in the lab, or in field experiments with small or unrepresentative groups. Furthermore, even if effective in field experiments, many of these academic interventions have been too labor-intensive to be used as a large-scale energy efficiency program.

Positive Energy intentionally designed its pilot programs to exploit this existing body of knowledge. While the randomization into treatment and control will allow an unbiased estimate of the overall effect of the Reports, because the entire Treatment group received the social comparisons, historical consumption information, and energy efficiency tips, it will not be possible to determine what aspect of the Reports drives that treatment effect. As in Benartzi and Thaler (2004), I can evaluate the overall efficacy of the program motivated by behavioral science, but I cannot test more refined hypotheses about the channels through which the program works. Although popular media outlets have concluded it is the peer comparison feedback in particular that reduces households' electricity usage, the treatment arms of the pilot program itself provide no evidence that this is actually true. What's interesting about the Positive Energy pilot experiments is that they take an existing body of scientific knowledge and implement it at scale.

2.2 Experimental Design

As of summer 2009, Positive Energy has contracts to operate pilot projects for utilities in California, Minnesota, Washington, Illinois, Colorado, and Virginia. This paper evaluates their program at an electric utility called Connexus Energy, which serves customers in seven counties in central Minnesota near Minneapolis and St. Paul[6].

The households eligible for the Connexus Energy pilot experiments were those with a full one year of electricity bill history as of January 2009, as historical consumption data were required to construct the social comparisons. From the 78,492 eligible households, approximately half (39,217) were randomized into a Treatment group, which would receive Home Energy Reports, and the rest

[6] Positive Energy carries out internal statistical evaluations, and independent evaluations of their pilot programs in Puget Sound and Sacramento are carried out by Ayres, Raseman, and Shih (2009) and Violette, Provencher, and Klos (2009).

were randomized into a Control group, which would not. Some utility staff were automatically enrolled in the reports and are therefore excluded from the analysis. The experiments are still ongoing.

The Reports are several-page letters with two key components. The first, which is illustrated in Figure 8.1 and appears at the top of the letter's first page, is the Social Comparison Module. This compares the household's electricity consumption over the past twelve months to the mean of its comparison group and the 20th percentile. In an attempt to address the potential boomerang effect, the "Efficiency Standing" on the right side of the Social Comparison Module adds injunctive messages: it labels low- and moderate-consumption households as "Great" and "Good" and adds "smiley face" emoticons.

The Report's second key component is the Action Steps Module. As illustrated in Figure 8.2, this suggests both changes to the household's stock of energy-using durable goods and to the use of that capital stock. These suggestions are targeted to different households based on historical energy use patterns and demographic characteristics. For example, households whose energy use was relatively high the previous summer were more likely to receive suggestions to purchase new energy efficient air conditioners.

The mechanics of the billing and report mailing process are as follows. The utility sends a worker to read each household's electricity meter approximately once per month, records the consumption over the period, and sends the consumer its monthly bill. Meanwhile, the meter readings are sent electronically to Positive Energy, where each household's social comparison is computed. The Home Energy Report is printed by an outside contractor, send via U.S. Mail, and at some point is opened at the household. The time between meter reading and the arrival of the Report is typically about three weeks. For almost all households, the first round of Reports were constructed after each household's meter reading in January 2009, although a small handful of households were randomized into Treatment and Control groups over the next few months. Sixty percent of Treatment group households were randomly assigned to receive the letters after every monthly meter reading, while 40 percent received them only once a quarter[7].

2.3 Data and Baseline Characteristics

I observe the 1.5 million electricity bills for all Treatment and Control households between January 2008 and August 2009, including the date of meter reading and consumption between that reading and the previous. I also observe Positive Energy's social comparison information for every household, including whether they were rated as "Below Average," "Good," or "Great," and how far they were from the cutoffs to be in each of the other categories. I observe both the social comparisons that the Treatment group did receive and what the Control group would have received. For each billing period, I also observe the number of Heating Degree-Days or Cooling Degree-Days, which are correlated with the amount of electricity that should be required to keep a house at a

[7] Two other variations in the treatment should be noted. After the second Report, the normative messaging was made more "Gentle" for a randomly selected half of the Treatment group. In the "Gentle" condition, the "Below Average" efficiency group no longer saw that there were "Great" and "Good" categories; they instead simply see the message that "You used more than average - Turn the report over to find ways to save." Positive Energy also experimented with sending different envelope types. Because the different envelope types and "Gentle" treatment do not have sharp economic interpretation, and because the estimated effects of these treatments are not substantially different, the treatment effects presented in this paper combine the different envelopes and the Gentle and non-Gentle treatments into one Average Treatment Effect.

comfortable temperature[8].

Table 7.1 displays the baseline observable characteristics for the Treatment and Control group. Baseline Usage, which is the average across all meter reads in 2008, is 29.7 kilowatt-hours per day for both the Treatment and Control groups. For households ever in the "Great," "Good," or "Below Average" groups, average Baseline Usage is 16.5, 25.0, and 40.7 kilowatt-hours per day, respectively. Although the household sizes are similar by construction, the "Great" group is poorer, has fewer household members, and has houses that are worth less.

For context, consider that a medium-sized (75 watt) lightbulb used four hours each day consumes 0.3 kilowatt-hours. A typical window air conditioner running at its highest setting for five hours uses 5 kilowatt-hours. As illustrated in Figure 8.3, heating and cooling are the primary uses of household electricity in the United States: over half of annual electricity consumption is for refrigerators, air conditioners, and space and water heating. In the most recent available data, computers, televisions, and lighting combined account for only 15 percent of electricity use (US Energy Information Administration 2001).

Connexus has provided demographic data for each account number in the dataset[9], including characteristics of the house (Age, an indicator for Gas Heat, Value, an indicator for Rental, Single Family, and Square Footage) and of the occupants (Age of household head, Household Size, and Income). On Baseline Usage, as well as on all other observable characteristics, the Treatment and Control groups are strikingly well-balanced. One of the ten baseline characteristics, the age of the head of household, is statistically different with 90% confidence; the Treatment and Control averages differ by less than 0.2 years. As would be expected in a randomized experiment, an F test fails to reject that the two groups are identical on observables.

2.4 Attrition

The pilot program experienced two forms of attrition, account closure and opting out. Consumers close accounts when they move to a different house; 1.25 and 1.24 percent of Treatment and Control households, respectively, close accounts during the first six months of 2009. Households that close accounts tend to be younger, use less electricity, and have lower incomes, but are statistically indistinguishable on other observed characteristics. There is no statistical difference between the Treatment and Control groups in either the rate of account closure or the correlations between account closure and observable characteristics. The accounts that closed after Treatment began are included in the base specifications during the period when their consumption is observed, although excluding them has no discernible influence on the results.

The second form of attrition is by households that asked to opt out of receiving the Home Energy Reports. In the first six months of 2009, 247 households (0.6 percent of the Treatment group) opted out, likely because they perceived the reports as undesired junk mail. These consumers are statistically different: they are slightly older, have lower incomes, use less electricity, and are less likely to live in single family homes. Although they opted out of receiving Reports, their electricity bills are still observed.

[8] More precisely, Heating Degree-Days is the sum, over all of the days in the billing period, of the maximum of zero and the difference between the day's average temperature and 65 degrees. A day with average temperature 95 has 30 HDDs, while a day with average temperature 60 has zero HDDs. Cooling Degree-Days is the sum, over all the days in the billing period, of the maximum of zero and the difference between 65 degrees and the day's average temperature. A day with average temperature 95 has zero CDDs, while a day with average temperature 60 has five CDDs.

[9] Any missing observations were imputed using conditional mean imputation.

3 Empirical Strategy

3.1 Average Treatment Effects

This section details the straightforward approach to estimate the Population Average Treatment Effect of the Home Energy Reports in the population of eligible households. Simply put, the preferred specification will estimate energy consumption as a function of whether the household is assigned to treatment, conditional on other controls, after removing household fixed effects.

To arrive at that specification, I begin with the standard Rubin Causal Model (Rubin 1974, Imbens and Wooldridge 2009). Each household i has two *potential outcomes* for the energy consumption outcome for the meter read t: one if it were assigned to the Treatment group ($T_i = 1$) and one if it were assigned to the Control ($T_i = 0$). Of course, only one of those outcomes can actually be observed:

$$Y_{it} = Y_{it}(T_i) = Y_{it}(0)(1 - T_i) + Y_{it}(1)(T_i) = \left\{ \begin{array}{ll} Y_{it}(0) & \text{if } T_i = 0 \\ Y_{it}(1) & \text{if } T_i = 1 \end{array} \right\} \tag{1}$$

The quantity of interest is the Average Treatment Effect, $\tau = E[Y_{it}(1) - Y_{it}(0)]$. The "treatment" here is defined as "being mailed the Home Energy Reports or actively opting out." As discussed above, some Treatment households opted out, so the treatment is not simply "being mailed the Home Energy Reports." An alternative potential estimand would be the Intent-to-Treat (ITT) Effect of being sent the Report, for the population that did not opt out; this is simply my ATE divided by the fraction of the population that did not opt out. Since that fraction is very close to 1, the ATE and ITT Effect differ only by a negligible amount. The treatment is also not "opening Home Energy Reports." It is difficult to measure letter open rates, and thus it would be difficult to estimate that second form of ITT Effect.

From a policy perspective, the treatment effect as I define it is the most useful estimand. Positive Energy, and the utilities that contract with them and policymakers that regulate them, want to know the aggregate electricity conservation possible from applying the program to a population. For the population from which the experimental households were drawn, this quantity of interest can be derived simply by multiplying the ATE by the population size[10].

Each household has a different meter reading schedule. Let $t \in \{t_{\min}, ..., t_{\max}\} = \{-12, -11, ..., 0, 1, 2, ...\}$ index bill numbers beginning one year before the treatment began; for each household, each bill number is associated with a particular day, month, and year. Most of the first set of Reports were sent in mid-January, immediately after a meter read at time t_{i0}. As I will show, little substantive or statistical effect is observed until the second meter reading after the January round. All meter reads t more than 40 days after t_{i0}, which for nearly all households is simply the second subsequent monthly meter read, are therefore considered "post-treatment."

The variable P_{it} is a post-treatment indicator variable for household i's meter reading at date t. Denote by Y_{it} the average daily electricity consumption for period ending at t. So that this can later be interpreted as a percentage change, the variable is normalized by control group consumption in the post-period. The variable Q_i denotes whether household i was assigned to the Quarterly group, in either Treatment or Control state. Random assignment to Treatment and Control implies that unobservable factors ε_{it} that influence electricity consumption are uncorrelated with T and Q.

[10] To determine cost-effectiveness, the cost estimates must be adjusted to account for the fact that some households will opt out.

This allows an unbiased estimate of the ATE for both the Monthly and Quarterly Groups with the following equation:

$$Y_{it} = (\tau + \tau_Q Q_i) \cdot T_i P_{it} + (\beta_1 + \beta_2 Q_i) \cdot P_{it} + \mu_{my} + \upsilon_i + \varepsilon_{it} \tag{2}$$

This specification includes month-by-year dummy variables μ_{my} and household fixed effects υ_i; alternative arrangements of fixed effects and controls will also be presented. This is estimated in OLS using the standard fixed effects estimator, using Huber-White ("robust") standard errors, clustered by household. As discussed by Bertrand, Duflo, and Mullainathan (2004), these standard errors are consistent in the presence of any correlation pattern in the errors ε_{it} within household over time.

3.2 Quarterly Group Effects Over Time

After estimating the Average Treatment Effects, I move to an empirical test of whether these effects decay in the Quarterly group over the three bills observed in each quarter. Intuitively, we would like to test whether the Quarterly group conserves more on the first and second bills after receiving a Report (the second and third bills after the one on which the report was based) compared to on the third bill of the quarter. This must control for bill-to-bill (i.e. month-to-month) variation in the average treatment effect, which could be driven by seasonal weather changes. I estimate the following equation:

$$\begin{aligned} Y_{it} = Q_i \cdot \{T_i P_{it} \cdot (\tau_Q + \tau_{Q12} B_{12it}) + P_{it} \cdot (\beta_1 + \beta_2 B_{12it})\} \\ + \sum_{b=t_{\min}}^{t_{\max}} 1(t=b) \cdot \{\beta_{1b} T_i + \beta_{2b}\} + \upsilon_i + \varepsilon_{it} \end{aligned} \tag{3}$$

The variable B_{12it} is an indicator for whether bill t is the first or second bill after receiving a Report. The first line compares the treatment effects τ_{Q12} in the Quarterly group when $B_{12it} = 1$ to the baseline quarterly treatment effect τ_Q. If the treatment effects decay, τ_{Q12} will be more strongly negative than τ_Q. The second line of the equation controls for underlying treatment effects and consumption levels for each bill number, ranging from $t_{\min} = -12$ at the beginning of the data to the most recent bill observation $t_{\max}$.

4 Results

4.1 Treatment Effects

Table 7.2 presents the estimates of the Average Treatment Effect in the eligible population. The top row is the ATE for the Monthly group, while the second row is the difference between that and the Quarterly group ATE. The five specifications include different configurations of fixed effects, month-by-year dummies, and weather controls; number III is the exact specification detailed in the Empirical Strategy section. The point estimates of ATEs center around negative 1.9 percent (Monthly) and 1.1 percent (Quarterly) and are not statistically different between the five specifications.

Table 7.3 displays two alternative approaches to evaluating these pilot programs, motivated by the fact that some in industry are unfamiliar with randomized evaluations and with the natural log function. The first two columns disregard the Control group and use a difference estimator to estimate the treatment effect. The first column is a simple before-after comparison, without accounting for weather or season, and the results are not encouraging: $\widehat{\tau}$ differs from the experimental estimate by almost an order of magnitude. After including fifth-order polynomials in Heating and Cooling Degree-Days and 12 month-of-year dummies in the second column, however, we have an estimated treatment effect that is statistically indistinguishable from and substantively similar to the experimental estimate. Although these specifications were initially included to demonstrate the importance of implementing the program as a randomized trial, it appears that in this particular case, the ATE could have been estimated even in the absence of a control group.

The third specification in Table 7.3 replicates Specification III from Table 7.2, except after logging the dependent variable. In principle, both coefficients are interpreted as percents and should be comparable. In practice, the estimated Monthly (and Quarterly) ATEs in the preferred specification are 42 (and 34) percent higher than that from the log specification. This difference is both substantively significant, in that it would make a noticeable difference in the estimated cost effectiveness, and statistically significant. This difference between the percentage ATEs in logs and levels is relatively constant across (unreported) different specifications, and it is not driven by the 307 observations of the dependent variable that are equal to zero.

Intuitively, this difference is driven by Jensen's Inequality. The treatment effect of interest is the average reduction in kilowatt-hours resulting from the program, and the percent changes reported in Table 7.2 are indeed that quantity, normalized by Control group consumption in the post-treatment period. The log specification is fundamentally different: it is the average percent reduction across households, where each household is effectively normalized to its own average (instead of the Control group average) before computing the ATE. It turns out that the percentage treatment effect is larger (in absolute value) for households with higher consumption, and substantially higher for the most consumptive households. Taking the log of the household's consumption, and then taking the average across households, will understate the ATE (in absolute value) relative to taking the average level change across households and then normalizing into a percent. Although this specification was initially included to demonstrate the innocuousness of using logs to compute percent changes, it is clear that in this case, that approach would be misleading.

Figure 8.4 illustrates the estimated Average Treatment Effect on meter reads over time, separately for the households receiving quarterly and monthly reports. The omitted meter read, number zero, is January for almost all households, the meter reading t_{i0} upon which the first reports were based. For the 12 reads before that, the treatment and control group consumption levels are substantively and statistically identical. This is still the case on the first meter read after t_{i0}. By the second meter read, however, there is a noticeable Average Treatment Effect. That affect grew between February and July, perhaps both because the response to the Reports grows in the early phases of the program and because the effects are stronger in the summer months.

Although the standard errors are wide, the figure suggests that the Quarterly group's ATE is higher in month 2, the first meter read where we would expect to see effects from a Report generated from the month 0 meter read, and then decays in months 3 and 4. A second quarterly report was generated from meter read number 3, and the Quarterly group's treatment effect again increases (in absolute value) in months 5 and 6.

Table 7.4 presents the formal econometric test, described in the Empirical Strategy section, of whether the effects are stronger for meter reads 2, 3, 5, and 6 relative to number 4. Specification I includes separate dummies for the first (and second) bill after receiving the report, which on the graph are reads 2 and 5 (and 3 and 6) respectively. The point estimate ATE for the first bill after receiving a Report is stronger than the point estimate for the second, although this difference is not close to statistically significant. Specification II is identical to Specification I except that it combines the dummy variables for the first and second bills. Both specifications show that the ATEs for the first and second bills after receiving a Report are stronger than the baseline Quarterly ATE, which captures the effect in the third bill of the quarter, with approximately 90 percent confidence in this two-sided test.

Three factors could explain this result. First, the information contained in the Reports - or gathered in response to the Reports - has value that varies by season. Positive Energy targets a set of between 100 and 200 tips to households, and while some have effects that would vary negligibly by season ("Buy an energy efficient refrigerator"), others are fundamentally seasonal ("Replace your heater" or "Cover your pool"). Second, perceiving oneself as a "frugal" consumer of energy could enter consumers' utility functions directly, and the marginal utility of conservation could depend on how recently a Report arrived. Third, the Reports could remind households of opportunities that they already knew about to conserve energy, and that reminder effect could decay over time. The second two channels would be similar to the "Two Steps Forward, One Step Back" story in consumer credit, where fees remind consumers to avoid triggering fees in the future, but that reminder effect decays over months (Agarwal, Driscoll, Gabaix, and Laibson 2006).

Positive Energy has collected surveys of its treatment group that provide some further evidence on this issue. Treatment group households in one of their other pilot programs were asked to self report what behaviors they had changed as a result of receiving the Home Energy Reports. Some of the changes were seasonal changes to household capital stock: weather stripping windows, improving insulation, servicing the air conditioner. Many of the most frequently reported changes, however, were habitual: turning off lights, unplugging electronics, adjusting thermostats, and closing drapes. The more likely it is that these all-season behavioral changes underlie the treatment effects, the more likely it is that the decays in the Quarterly group ATEs are part of a cycle in which the Reports remind or motivate households to conserve, and this attention or motivation decays over months.

Figure 8.5 illustrates the treatment effects by deciles of the distribution of baseline usage, again normalized by Control group average consumption in the post-treatment period. These effects range from less than 0.5 percent for the bottom two deciles of baseline usage to 6.2 percent in the top ten percent. In general, the more electricity a household used before the treatment, the more that it conserved post-treatment. This could be because the most consumptive households had low-cost energy conservation opportunities, and the tips contained in the Reports made them aware of this. This result is also consistent with the idea that previously low-consumption households might not conserve more - or might even conserve less - after receiving information that they are less consumptive than their peers.

4.2 Regression Discontinuity

Some readers may have noticed that the design of the normative categorizations could allow the use of a Regression Discontinuity design to estimate their causal relative effects. Those households just below the cutoff between being categorized "Good" and "Great" are in the limit identical to those households just above, but they received different normative categorizations. This provides a natural experiment that could allow the estimation of the relative effects, for households near the cutoff, of being in the three different categorizations. These effects should be observed two meter reads after the meter read upon which the categorization was based.

In keeping with the tradition of graphical analysis of RD designs (Lee and Lemieux 2009, Imbens and Wooldridge 2009), Figure 8.6 illustrates the treatment group's consumption as a function of distance from the household-specific comparison group's mean consumption, which is the cutoff between being categorized as "Good" vs. "Below Average." The Usage variable on the y-axis, which as before is normalized by Control group post-treatment consumption, is residual of degree-day polynomials and month-by-year controls, but not of household fixed effects, so those households that consume less compared to their peers also tend to have lower residual usage on future bills. Figure 8.7 is the analogous illustration near the 20th percentile of the household-specific comparison group, which is the cutoff between being categorized as "Great" vs. "Good."

What's apparent from both graphs is that the two-sided 95 percent confidence interval around any estimated effects at this bandwidth is approximately four percent - more than double the overall ATE from the Reports. Any differences in ATEs from being placed in one or the other normative category, however, would likely be much smaller. I attempted the RD estimation including household fixed effects and other controls to reduce residual variance, and also added data from one of Positive Energy's other pilot programs. This does not sufficiently reduce the variance around the estimated RD treatment effect to estimate a statistically significant effect of the categorizations or to compute a "tightly-estimated zero" by rejecting reasonably small hypothesized effects.

5 Cost Effectiveness and Profiling

Regardless of the mechanism that drives the variation in treatment effects across households with different baseline usage, the heterogeneity in treatment effects as a function of an observable characteristic suggests that there could be substantial gains from targeting the program towards the most responsive households. Furthermore, because we observe a larger set of household characteristics that might be correlated with the treatment effect, a more comprehensive approach to "profiling" could be useful. My approach builds on the literature detailing the use of existing information on heterogeneous treatment effects to allocate future treatments. This literature dates to Wald's (1950) work on statistical decision theory and has grown recently with work by Berger, Black, and Smith (2000), Dehejia (2005), Graham, Imbens, and Ridder (2009), Hirano and Porter (2006), Imai and Strauss (2009), and Manski (2004, 2009).

Compared to much of the recent literature, this problem is straightforward, as the objective function will be unambiguous and the decisionmaker can be modeled as risk-neutral. As discussed in the introduction, the typical reason why utilities contract with Positive Energy is to help comply with state energy efficiency portfolio standards. I model the decisionmaker as a utility that is trying to minimize its expected cost of compliance with this regulation, which requires a reduction in consumption of S kilowatt-hours from the counterfactual baseline[11]. The key decision variable is which subset of its customer population to assign to Positive Energy treatment.

[11] This requires the simplifying assumption that the regulator allows the utility to comply in expectation as opposed to *ex post*. These regulations often allow the utility to "bank" and "borrow" energy savings across years, meaning that in practice, the utility needs to comply *ex post* only over a long time period.

While we would in principle want to maximize welfare, the utility in practice minimizes cost from its own perspective, with no consideration for the change in consumer welfare. In designing this system, regulators presumably hope that the solutions to the two distinct optimization problems are similar[12]. In practice, and in this analysis in particular, this simplification must be made because while it is easy to observe electricity consumption, it would be quite costly to observe the actions that households take in response to the Reports. It is possible, for example, that the Reports induce some households to buy more energy efficient lightbulbs or appliances. Since the change in treatment group demand for these other goods is unobserved, we cannot compute welfare. This concern is quite general in evaluating energy efficiency programs; one resulting benefit here is that focusing on cost effectiveness keeps this analysis consistent with a previous body of work.

Positive Energy costs the utility a constant amount A per household per year. The utility also has a portfolio of other energy efficiency programs that it knows with certainty will cost C_0 per kilowatt hour, and it sets the size of these programs $K > 0$ in kilowatt-hours saved per year. The decisionmaker has information Θ from the existing large randomized experiment in a representative sample of larger population $\mathcal{P}$.

The decisionmaker seeks a statistical treatment rule $\delta : \mathcal{X} \longrightarrow \{0, 1\}$ that maps individuals with characteristics X to the treated or control states for the Positive Energy program. Denote by Δ the space of possible treatment rules. The scalar $\delta_i = \delta(X_i) \in \{0, 1\}$ is the choice of treatment for individual i with characteristics X_i. The heterogeneous Conditional Average Treatment Effect $\tau(X_i)$ is still typically less than zero, but now consider the units to be in kilowatt-hours per year.

The utility's objective is to minimize C, the expected annual cost of compliance with the regulation:

$$\begin{aligned} \min_{\delta \in \Delta, K} C &= \sum_{i \in \mathcal{P}} \delta_i \cdot A + K \cdot C_0 \\ s.t. \quad E\left[\sum_{i \in \mathcal{P}} -\tau(X_i) \cdot \delta_i | \Theta\right] + K &\geq S \end{aligned}$$

The utility's optimal statistical decision rule collapses to assigning Positive Energy treatment to those households where cost effectiveness is above some threshold value. Given that the cost A is constant across households, this can be written as a constraint on the expected treatment effect:

$$\delta_i^* = 1\left(E\left[-\tau(X_i)|\Theta\right] > R^*\right) = 1\left(\widehat{-\tau(X_i)} \geq R^*\right)$$

$$R^* = \left\{ \begin{array}{ll} \min\{\delta_i^* \tau_i\} & s.t. \sum_{i \in \mathcal{P}} \delta_i^* \cdot E\left[-\tau(X_i)|\Theta\right] = S, \quad K^* = 0 \\ & \frac{A}{C_0}, \qquad K^* > 0 \end{array} \right\}$$

[12] One example of why this distinction can be important is the case of rebates for energy efficient appliances. If these rebates are entirely inframarginal (an example of what energy industry analysts call the "free rider problem," not to be confused with the traditional free rider problem in the provision of public goods), their cost effectiveness would be infinite. Given that inframarginal rebates are simply transfers, the efficiency losses would be zero.

The threshold value R^* depends on whether Positive Energy is in the optimum the only program used to satisfy the regulation, or whether other energy efficiency programs are also used and $K^* > 0$. The former case is captured by the second line of the above equation: the utility needs only assign enough households to Positive Energy such that the constraint is satisfied, and there will be some threshold household with lowest cost-effectiveness. The latter case is captured in the third line: if other programs are used, the utility will assign to Treatment all households whose expected cost effectiveness is above the cost effectiveness of the alternative programs.

While recent related applications such as Dehejia (2005) and Imai and Strauss (2009) have used Bayesian estimators to construct a distribution of possible outcomes, I continue with the frequentist approach more familiar to program evaluation in economics. As suggested in the first line of the above equation, I simply use the heterogeneous treatment effects estimated in OLS, $\widehat{\tau_i(X_i)}$ to construct $E[\tau(X_i)|\Theta]$. Mechanically, I then generate that fitted treatment effect for each household, and treatment can be assigned for households above any given R.

5.1 Variable Selection

Estimating $\widehat{\tau_i(X_i)}$ requires the selection of a vector of observables and interactions X to condition on. As discussed in Imai and Strauss (2009), this is a critical decision. Including a large set of controls could allow treatment to be targeted at smaller subgroups that might have particularly large treatment effects. In a finite sample, however, including a larger set of covariates increases the likelihood of overfitting, which could cause the program to be targeted at groups that idiosyncratically appeared to have large treatment effects only in the experimental data.

Imai and Straus (2009) point out that while there is a large literature on how to select "predictive" variables that are correlated with the outcome, there is no consensus on the choice of "prescriptive variables" that help assign units to treatment. Predictive variables are correlated with the outcome, but if they have the same correlation in treatment and control, they are useful for reducing residual variance (improving efficiency), but not for generating a statistical decision rule. Powerful prescriptive variables have two features. First, they are correlated with the treatment effect. Second, they affect the decisions made for each unit.

Future versions of this analysis will use the approach of Zhu, Gunter, and Murphy (2007) to choosing the set of variables to interact with the treatment effect. This version uses three simple covariate specifications, all of which are linear interactions of the treatment effect τ in Equation I with a vector of observable covariates X, such that $\tau = \tau_0 + \tau_X X$. The first specification interacts the treatment indicator variable with all nine available characteristics. The second interacts the treatment indicator only with the three characteristics on which the treatment effect varies with 95 percent confidence in the first specification. The third interaction uses only Baseline Usage.

Table 7.5 presents the results of these specifications. The covariates are normalized to mean zero, standard deviation 1, meaning that a one standard deviation increase in Baseline Usage is associated with an increase in the ATE (in absolute value) of 1.50 to 1.191 percentage points. Houses with gas heat have higher treatment effects in absolute value, while houses with larger square footage have lower treatment effects. No other observable characteristics are statistically significantly associated with the strength of the treatment effect.

5.2 Gains from Profiling

Table 7.6 presents the predicted ATEs for assigning treatment to one half of the population based on the three sets of profiling X variables, compared to assigning to treatment the entire population or a randomly selected sample thereof. These predicted ATEs are computed by re-estimating the treatment effects with the treatment and control observations from the original experiment that now have $\delta = 1$. By this measure, profiling increases the ATE for the treated group by 80%. Interestingly, although the three different sets of X variables assign only 70 percent of units to the same treatment status, the resulting treatment effects on the treated populations are very similar across the three profiling procedures.

The table's second row displays the average annual electricity bill savings for the treated group under the four different assignment mechanisms, based on Connexus Energy's average residential electricity price[13]. The average household saves $15 per year on electricity bills as a result of the changes caused by being sent monthly Home Energy Reports. The table's third row presents the monthly program's cost effectiveness, based on my estimate of the cost per household of administering monthly Reports. The cost effectiveness of the existing randomized program is 5.3 cents per kilowatt-hour saved[14]. This compares very favorably to recent point estimates[15] of the average cost of other utility energy efficiency programs, which range from 4.7 to 13.3 cents per kilowatt-hour (Auffhammer, Blumstein, and Fowlie 2008). Using profiling to target monthly[16] treatment at the most responsive 1/2 of the population improves cost effectiveness by 45 percent, to below 3 cents-kilowatt-hour[17].

6 Conclusion

This paper evaluates the effects of the Positive Energy Home Energy Reports, which give households easily-understandable feedback on past energy consumption, compare them to their neighbors, and provide energy conservation tips. The program is a remarkable departure from traditional energy efficiency programs in that it is designed with direct insight from behavioral science and is implemented using randomized controlled trials. The perceived success or failure of these pilot programs will directly affect millions of dollars of future investment and will more generally influence how future energy efficiency programs are designed and evaluated.

I find that the Average Treatment Effect in the population of eligible Minnesota households is 1.9 percent below baseline. In the group receiving quarterly Reports, his effect decays somewhat

[13] This is available from http://www.connexusenergy.com/resrates.htm.

[14] Note that this cost effectiveness value is different than the number in Allcott and Mullainathan (2009a, 2009b), which is for a large-scale implementation across the country. The appropriate assumptions for baseline electricity demand and program cost differ between the two settings.

[15] I write "point estimate" because there are various forms of uncertainty around this comparison, including not just the statistical uncertainty in the point estimate conditional on the data, but also uncertainties such as regarding the durability of energy savings, the generalizability across different regions, and the quality of the underlying data used to compute the costs of these other programs.

[16] The ATE of the Quarterly reports is much more than 1/3 the ATE of the monthly reports, and depending on Positive Energy's ratio of fixed to variable costs, Quarterly reports could be more cost-effective.

[17] Note that this example is not a reasonable basis for an argument to limit the application of the program: given the low cost of Positive Energy's program, even applying the program households where the treatment effect is smaller than the median can be more cost effective than existing alternative programs. This does, however, show that if the program must be limited to some subset of the population due to budget constraints, there are gains to targeting it towards the most responsive subset.

in the months between Reports, either because the information decays seasonally or because the reminder or motivational effects of a Report decay over time. I also show that the intervention's effects are strongest for households that have highest baseline consumption. This is consistent with (but not causal evidence of) a "boomerang effect" in which learning the social norm can fail to motivate households with low baseline consumption, or even cause them to increase consumption. This is also consistent with the idea that the energy conservation tips in the Reports were more useful for high-consumption households, independent of the social norms.

I then build on the idea of heterogeneous treatment effects to design a statistical decision rule to target the program at households with the highest expected treatment effects. If a utility were limited to implementing the program for half of its residential population, profiling could improve cost-effectiveness by 80%. Even without profiling, the cost effectiveness of the Positive Energy program compares very favorably to traditional energy efficiency programs, which are largely based on expensive subsidies for energy efficient durable goods.

Although the Positive Energy experiment was carried out in a specific domain and requires a general-interest reader to digest some institutional detail, this analysis has two important and generalizable economic implications. First, the analysis adds to a recently-growing appreciation of how "profiling," the targeting of social programs with heterogeneous treatment effects toward individuals with high expected effects, can improve a program's welfare implications. Dehejia (2005) shows that a program that would not be implemented based on examining the ATE alone would optimally be implemented in subgroups with higher Conditional Average Treatment Effects. The Positive Energy example is different in that the intervention is cost effective relative to other energy efficiency programs even in the general population, but I similarly conclude that profiling could substantially increase the intervention's average cost effectiveness.

A second important implication of this analysis is that it adds to recently-growing appreciation of how non-price interventions can affect consumer behavior. Economists in general, and energy efficiency program managers in particular, have historically focused on how prices and subsidies affect demand. The idea that simply being sent a letter in the mail could result in measureable changes in demand is remarkable, especially given that the letters may not have improved consumers' information sets in a relevant way: recall that survey evidence indicates that the bulk of behavioral changes caused by the program are behaviors such as turning off lights that consumers *already knew* could save them energy. From the utility's perspective, interventions such as this are remarkably cheap to implement relative to manipulating prices, and comparable treatment effects therefore implies much better cost effectiveness. Perhaps the most important contribution of this program, therefore, is to provide further evidence that some combination of information, reminders, and social norms can cause substantive changes in consumer behavior at population scale.

References

[1] Abrahamse, Wokje, Linda Steg, Charles Vlek, and Talib Rothengatter (2005). "A Review of Intervention Studies Aimed at Household Energy Conservation." Journal of Environmental Psychology, Vol. 25, No. 3 (September), pages 273-291.

[2] Agarwal, Sumit, John Driscoll, Xavier Gabaix, and David Laibson (2006). "Two Steps Forward, One Step Back: The Dynamics of Learning and Backsliding." Working Paper, Harvard University (July).

[3] Aigner, Dennis (1984). "The Welfare Econometrics of Peak-Load Pricing for Electricity." *Journal of Econometrics*, Vol. 26, No. 1-2, pages 1-15.

[4] Akerlof, George (1980). "A Theory of Social Custom, of which Unemployment May Be One Consequence." *Quarterly Journal of Economics*, Vol. 94, No. 4 (June), pages 749-775.

[5] Akerlof, George, and Rachel Kranton (2000). "Economics and Identity." Quarterly Journal of Economics, Vol. 115, No. 3 (August), pages 715-753.

[6] Allcott, Hunt (2009). "The Effects of Residential Real Time Pricing: Evidence from a Randomized Trial." Working Paper, Massachusetts Institute of Technology (July).

[7] Allcott, Hunt, and Sendhil Mullainathan (2009a). "Behavioral Science and Energy Conservation." Working Paper, Massachusetts Institute of Technology (July).

[8] Allcott, Hunt, and Sendhil Mullainathan (2009b). "Behavioral Science and Energy Policy." Working Paper, Massachusetts Institute of Technology (August).

[9] Alpizar, Francisco, Fredrik Carlsson, and Olof Johansson-Stenman (2008). "Anonymity, Reciprocity, and Conformity: Evidence from Voluntary Contributions to a National Park in Costa Rica." *Journal of Public Economics*, Vol. 92, pages 1047-1060.

[10] Andreoni, James, and Ragan Petrie (2004). "Public Goods Experiments Without Confidentiality: A Glimpse Into Fund-Raising." *Journal of Public Economics*, Vol. 88, pages 1605-1623.

[11] Ashraf, Nava, Dean Karlan, and Wesley Yin (2006). "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines." *Quarterly Journal of Economics*, Vol. 121, No. 2, pages 673-697.

[12] Aubin, Christophe, Denis Fougere, Emmanuel Husson, and Marc Ivaldi (1995). "Real-Time Pricing of Electricity for Residential Customers: Econometric Analysis of an Experiment." *Journal of Applied Econometrics*, Vol. 10 (December), pages S171-S191.

[13] Banerjee, Abhijit (1992). "A Simple Model of Herd Behavior." *Quarterly Journal of Economics*, Vol. 107, No. 3 (August), pages 797-817.

[14] Barbose, Galen, Charles Goldman, and Bernie Neenan (2004). "A Survey of Utility Experience with real time pricing." Working Paper, Lawrence Berkeley National Laboratory, December.

[15] Becker, Gary (1965). "A Theory on the Allocation of Time." *Economic Journal*, Vol. 75, pages 493-517.

[16] Becker, Gary (1974). "A Theory of Social Interactions." *Journal of Political Economy*, Vol. 82, No. 6 (November), pages 1063-1093.

[17] Benabou, Roland and Jean Tirole (2003). "Intrinsic and Extrinsic Motivations." *Review of Economic Studies*, Vol. 70, pages 489-520.

[18] Benartzi, Schlomo, and Richard Thaler (2004). "Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving." *Journal of Political Economy*, Vol. 112, No. 1 (February), pages S164-S187.

[19] Berger, Mark, Dan Black, and Jeffrey Smith (2000). "Evaluating Profiling as a Means of Allocating Government Services." Working Paper, Syracuse University (September).

[20] Bernheim, Douglas (1994). "A Theory of Conformity." *Journal of Political Economy*, Vol. 102, No. 5(October), pages 847-877.

[21] Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004). "How Much Should We Trust Difference-in-Differences Estimates?" *Quarterly Journal of Economics*, Vol. 119, No. 1, pages 249-275.

[22] Bertrand, Marianne, Dean Karlan, Sendhil Mullainathan, Eldar Shafir, and Jonathan Zinman (2010). "What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment." *Quarterly Journal of Economics*, forthcoming.

[23] Beshears, John, James Choi, David Laibson, Brigitte Madrian, and Katherine Milkman (2009). "The Effect of Providing Peer Information on Retirement Savings Decisions." Working Paper, Harvard University (March).

[24] Bitler, Marianne, Jonah Gelbach, and Hilary Hoynes (2006). "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments." *American Economic Review*, Vol. 96, No. 4 (September), pages 988-1012.

[25] Blumstein, Carl, Seymour Goldstone, and Loren Lutzenheiser (2000). "A Theory-Based Approach to Market Transformation." Energy Policy, Vol. 28, No. 2 (February), pages 137-144.

[26] Boisvert, Richard N., Peter Cappers, Charles Goldman, Bernie Neenan, and Nicole Hopper (2007). "Customer Response to RTP in Competitive Markets: A Study of Niagara Mohawk's Standard Offer Tariff." *The Energy Journal*, Vol. 28, No. 1 (January), pages 53-74.

[27] Cialdini, Robert (2003). "Crafting Normative Messages to Protect the Environment." *Current Directions in Psychological Science*, Vol. 12, pages 105-109.

[28] Cialdini, Robert, Linda Demaine, Brad Sagarin, Daniel Barrett, Kelton Rhoads, and Patricia Winter (2006). "Managing Social Norms for Persuasive Impact." *Social Influence*, Vol. 1, pages 3-15.

[29] Cialdini, Robert, Raymond Reno, and Carl Kallgren (1990). "A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places." *Journal of Personality and Social Psychology*, Vol. 58, pages 1015-1026.

[30] Clee, Mona, and Robert Wicklund (1980). "Consumer Behavior and Psychological Reactance." *Journal of Consumer Research*, Vol. 6, pages 389-405.

[31] Conley, Timothy and Christopher Udry (Forthcoming). "Learning About a New Technology: Pineapple in Ghana." *American Economic Review.*

[32] Darby, Sarah (2006). "The Effectiveness of Feedback on Energy Consumption." Working Paper, Oxford Environmental Change Institute (April).

[33] Davis, Lucas (2008). "Durable Goods and Residential Demand for Energy and Water: Evidence from a Field Trial." *RAND Journal of Economics*, Vol. 39, No. 2 (Summer), pages 530-546.

[34] Dehejia, Rajeev (2005). "Program Evaluation as a Decision Problem." *Journal of Econometrics*, Vol. 125, pages 141-173.

[35] Duflo, Esther, and Emmanuel Saez (2002). "Participation and Investment Decisions in a Retirement Plan: The Influence of Colleagues' Choices." *Journal of Public Economics*, Vol. 85, pages 121-148.

[36] Duflo, Esther, and Emmanuel Saez (2003). "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment." *Quarterly Journal of Economics*, Vol. 118, No. 3 (August), pages 815-842.

[37] Fischbacher, Urs, Simon Gachter, and Ernst Fehr (2001). "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." *Economic Letters*, Vol. 71, pages 397-404.

[38] Foster, Andrew, and Mark Rosenzweig (1995). "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *Journal of Political Economy*, Vol. 103, No. 6 (December), pages 1176-1209.

[39] Frey, Bruno, and Felix Oberholzer-Gee (1997). "The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out." *American Economic Review*, Vol. 87, No. 4 (September), pages 746-755.

[40] Frey, Bruno, and Stephan Meier (2004). "Social Comparisons and Pro-Social Behavior: Testing 'Conditional Cooperation' in a Field Experiment." *American Economic Review*, Vol. 94, No. 5 (December), pages 1717-1722.

[41] Gerber, Alan, and Todd Rogers (2009). "Descriptive Social Norms and Motivation to Vote: Everybody's Voting and So Should You." *Journal of Politics*, Vol. 71, pages 1-14.

[42] Gillingham, Kenneth, Richard Newell, and Karen Palmer (2006). "Energy Efficiency Policies: A Retrospective Examination." *Annual Review of Environment and Resources*, Vol. 31, pages 161-192.

[43] Goldstein, Noah, Robert Cialdini, and Vladas Griskevicius (2008). "A Room with a Viewpoint: Using Norm-Based Appeals to Motivate Conservation Behaviors in a Hotel Setting." Journal of Consumer Research, Vol. 35, pages 472-482.

[44] Graham, Bryan, Guido Imbens, and Geert Ridder (2009). "Complementarity and Aggregate Implications of Assortative Matching: A Nonparametric Analysis." NBER Working Paper 14860 (April).

[45] Grinblatt, Mark, Matti Keloharju, and Seppo Ikäheimo (2008). "Social Influence and Consumption: Evidence From the Automobile Purchases of Neighbors." *Review of Economics and Statistics*, Vol. 90, pages 735-753.

[46] Gunter, Lacey, Ji Zhu, and Susan Murphy (2007). "Variable Selection for Optimal Decision Making." Artificial Intelligence in Medicine, Vol. 4594 (August), pages 149-154.

[47] Graham, Bryan, Guido Imbens, and Geert Ridder (2009). "Complementarity and Aggregate Implications of Assortative Matching: A Nonparametric Analysis." NBER Working Paper 14860 (April).

[48] Hirano, Keisuke, and Jack Porter (2006). "Asymptotics for Statistical Treatment Rules." Working Paper, University of Wisconsin (August).

[49] Houwelingen, Jeannet, and Fred van Raaij (1989). "The Effect of Goal-Setting and Daily Electronic Feedback on In-Home Energy Use." Journal of Consumer Research, Vol. 16, No. 1 (June), pages 98-105.

[50] Hutton, Bruce, Gary Mauser, Pierre Filiatrault, and Olli Ahtola (1986). "Effects of Cost-Related Feedback on Consumer Knowledge and Consumption Behavioral: A Field Experimental Approach." *Journal of Consumer Research*, Vol. 13, No. 3 (December), pages 327-336.

[51] Imai, Kosuke, and Aaron Strauss (2009). "Planning the Optimal Get-out-the-vote Campaign Using Randomized Field Experiments." Working Paper, Princeton University (May).

[52] Imbens, Guido, and Jeffrey Wooldridge (2009). "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature*, Vol. 47, No. 1 (March), pages 5-86.

[53] Imbens, Guido, and Karthik Kalyanaraman (2009). "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." IZA Discussion paper No. 3995 (February).

[54] Kaufman, Leslie (2009). "Utilities Turn Their Customers Green, With Envy." *The New York Times*, January 30.

[55] Lee, David, and Thomas Lemieux (2009). "Regression Discontinuity Designs in Economics." NBER Working Paper 14723 (February).

[56] Levitt, Steven D. and John A. List (2009). "Field Experiments in Economics: The Past, the Present, and the Future." *European Economic Review*, forthcoming.

[57] Levy, Sidney (1959), "Symbols for Sale." *Harvard Business Review*, Vol. 37 (July-August), pages 117-124.

[58] List, John, and Michael Price (2008). "The Role of Social Connections in Charitable Fundraising: Evidence from a Natural Field Experiment." *Journal of Economic Behavior and Organization*, Vol. 69, No. 2 (February), pages 160-169.

[59] Lutzenhiser, Loren (1993). "Social and Behavioral Aspects of Energy Use." Annual Review of Energy and the Environment, Vol. 18 (November), pages 247-289.

[60] Manski, Charles (2004). "Statistical Treatment Rules for Heterogeneous Populations." *Econometrica*, Vol. 72, No. 4 (July), pages 1221-1246.

[61] Manski, Charles (2009). "Diversified Treatment Under Ambiguity." Working Paper, Northwestern University.

[62] Meer, Jonathan (2009). "Brother Can You Spare a Dime? Peer Effects in Charitable Solicitation." Working Paper, Stanford Institute for Economic Policy Research (March).

[63] Mobius, Markus, Paul Niehaus, and Tanya Rosenblat (2005). "Social Learning and Consumer Demand." Working Paper, Harvard University (December).

[64] Munshi, Kaivan, 2004. "Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution." *Journal of Development Economics*, Vol. 73, pages 185-215.

[65] Munshi, Kaivan, and Jacques Myaux (2006). "Social Norms and the Fertility Transition." *Journal of Development Economics*, Vol. 80, pages 1-38.

[66] Nadel, Steven, and Kenneth Keating (1991). "Engineering Estimates versus Impact Evaluation Results: How Do They Compare and Why?" in Energy Program Evaluation: Uses, Methods, and Results. Proceedings of the 1991 International Energy Program Evaluation Conference.

[67] Nolan, Jessica, Wesley Schultz, Robert Cialdini, Noah Goldstein, and Vladas Griskevicius (2008). "Normative Influence is Underdetected." *Personality and Social Psychology Bulletin*, Vol. 34, pages 913-923.

[68] Pesendorfer, Wolfgang (1995). "Design Innovation and Fashion Cycles." *American Economic Review*, Vol. 85, No. 4 (September), pages 771-792.

[69] Reiss, Peter, and Matthew White (2005). "Household Electricity Demand, Revisited." *Review of Economic Studies*, Vol. 72, No. 3 (July), pages 853-883.

[70] Reiss, Peter, and Matthew White (2008). "What Changes Energy Consumption? Prices and Public Pressure." *RAND Journal of Economics*, Vol. 39, No. 3 (Autumn), pages 636-663.

[71] Ringold, Debra Jones (2002). "Boomerang Effects in Response to Public Health Interventions: Some Unintended Consequences in the Alcoholic Beverage Market." Journal of Consumer Policy, Vol. 25, No. 1 (March), pages 27-63.

[72] Rubin, Donald (1974). "Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies." *Journal of Educational Psychology*, Vol. 66, No. 5, pages 688-701.

[73] Schultz, Wesley, Jessica Nolan, Robert Cialdini, Noah Goldstein, and Vladas Griskevicius (2007). "The Constructive, Destructive, and Reconstructive Power of Social Norms." *Psychological Science*, Vol. 18, pages 429 –434.

[74] Shang, Jen, and Rachel Croson (2004). "Field Experiments in Charitable Contribution: The Impact of Social Influence on the Voluntary Provision of Public Goods." Working Paper, University of Pennsylvania.

[75] Shippee, Glenn (1980). "Energy Consumption and Conservation Psychology: A Review and Conceptual Analysis." *Environmental Management*, Vol. 4, No. 4, pages 297-314.

[76] Stern, Paul (1992). "What Psychology Knows about Energy Conservation." *American Psychologist*, Vol. 47, No. 10, pages 1224-1232.

[77] Tsui, Bonnie (2009). "Greening With Envy." *The Atlantic*, Vol. 304, No. 1 (July/August), pages 24-25.

[78] US Energy Information Administration (2001). "Residential Energy Consumption Survey." http://www.eia.doe.gov/emeu/recs/recs2001.

[79] Veblen, Thorstein (1899). The Theory of the Leisure Class: an economic study of institutions. New York, NY: The Macmillan Company.

[80] Violette, Daniel, Provencher, Bill, and Mary Klos (2009). "Impact Evaluation of Positive Energy SMUD Pilot Study." Boulder, CO: Summit Blue Consulting (May).

[81] Wald, Abraham (1950). Statistical Decision Functions. New York, NY: John Wiley & Sons.

[82] Wernerfelt, Birger (1990). "Advertising Content When Brand Choice Is a Signal." *Journal of Business*, Vol. 63, No. 1, Part 1 (January), pages 91-98.

[83] Wolak, Frank (2006). "Residential Customer Response to Real-time Pricing: The Anaheim Critical Peak Pricing Experiment." Center for the Study of Energy Markets Working Paper 151 (May).

7 Tables

7.1 Baseline Household Characteristics

	Treatment	*Control*	*T-C*	*Great*	*Good*	*Below Av*
Mean: Baseline Usage (kwh/day)	29.74	29.69	0.053	16.52	25.04	40.67
SD (and SE):	(16.44)	(16.17)	(0.117)	(8.05)	(9.68)	(17.89)
Consumer Age	49.98	49.82	0.161	52.57	50.05	48.45
	(12.18)	(12.18)	(0.087)*	(12.99)	(12.34)	(11.38)
1(Gas Heat)	0.92	0.92	-0.0013	0.94	0.93	0.90
	(0.28)	(0.27)	(0.0020)	(0.24)	(0.25)	(0.30)
Household Size	2.61	2.62	-0.010	2.23	2.57	2.84
	(1.22)	(1.23)	(0.009)	(1.07)	(1.18)	(1.28)
House Age	16.27	16.21	0.062	16.56	16.27	16.02
	(6.65)	(6.56)	(0.047)	(6.56)	(6.59)	(6.64)
House Value	393.9	394.7	-0.792	369.6	385.5	414.5
	(139.4)	(139.4)	(0.995)	(121.7)	(132.6)	(150.6)
Income (1000s)	86.18	86.19	-0.010	75.59	84.42	92.85
	(37.86)	(37.60)	(0.269)	(33.52)	(36.48)	(39.52)
1(Rent)	0.022	0.023	0.000	0.019	0.023	0.024
	(0.14)	(0.14)	(0.001)	(0.11)	(0.14)	(0.15)
Single Family	0.96	1	-0.001	0.97	0.96	0.95
	(0.21)	(0)	(0.001)	(0.18)	(0.20)	(0.22)
Square Footage	1660	1658	1.68	1615	1637	1702
	(454)	(447)	(3.22)	(440)	(440)	(463)
F-Test p-Value			0.315			
1(Account Closed)	0.012	0.013		0.023	0.011	0.010
	0.11	0.11		0.15	0.10	0.10
Number of Households	39217	39275		14,055	33,016	30,913
Number of Bill Obs	726954	728623		261,721	614,809	576,387

*, **, ***: Different from zero with 90%, 95%, and 99% confidence, respectively.

"Number of Households" by normative categorization reflects the number of treatment or control households that were at any point in that category. The sum of these numbers across the three categories therefore is greater than the number of households in the experiment.

7.2 Treatment Effects

	I	II	III	IV	V
T x Post	-0.0175	-0.0192	-0.0191	-0.0208	-0.0189
	(0.0053)***	(0.0020)***	(0.0020)***	(0.0021)***	(0.0020)***
T x Quarterly x Post	0.0074	0.0085	0.0086	0.0080	0.0082
	(0.0082)	(0.0032)***	(0.0032)***	(0.0033)**	(0.0032)***
Post	-0.1167	-0.1135	-0.0465	-0.0873	-0.0062
	(0.0033)***	(0.0014)***	(0.0059)***	(0.0052)***	(0.0058)
Quarterly x Post	-0.0099	-0.0046	-0.0047	-0.0058	-0.0044
	(0.0058)*	(0.0022)**	(0.0022)**	(0.0023)**	(0.0022)**
Degree-Day Polynomial					Yes
Month x Year Dummies			Yes	Yes	Yes
House Fixed Effects		Yes	Yes		Yes
House x Month Fixed Effects				Yes	
Observations (thousands)	1,456	1,456	1,456	1,456	1,456
R^2	0.0057	0.0057	0.0472	0.0018	0.0499
F Statistic	5406	6412	7356	929	7188

*, **, ***: Different from zero with 90%, 95%, and 99% confidence, respectively.

Dependent variable is the household's average daily electricity consumption (kilowatt-hours), normalized by average control group consumption in the Post period.

7.3 Alternative Empirical Approaches

	Treatment Only	Treatment Only	In Logs
T x Post	-0.1327	-0.0206	-0.0134
	(0.0015)***	(0.0018)***	(0.0017)***
T x Quarterly x Post	0.0039	0.0038	0.0064
	(0.0023)*	(0.0023)	(0.0029)**
Post			-0.0413
			(0.0039)***
Quarterly x Post			-0.0062
			(0.0020)***
Degree-Day Polynomial		Yes	
Month Dummies		Yes	
Month x Year Dummies			Yes
House Fixed Effects	Yes	Yes	Yes
Observations (thousands)	727	727	1455
R^2	0.0189	0.0503	0.0579
F Statistic	6744	908	14043

*, **, ***: Different from zero with 90%, 95%, and 99% confidence, respectively.

Dependent variable for the first two specifications is the household's average daily electricity consumption (kilowatt-hours), normalized by average control group consumption in the Post period. For the third specification, it is the log of that value.

7.4 Decay of Quarterly Treatment Effects

	I	II
T x Quarterly x Post	0.0140	0.0140
	(0.0049)***	(0.0049)***
T x Quarterly x Post (1st Bill)	-0.0081	
	(0.0048)*	
T x Quarterly x Post (2nd Bill)	-0.0067	
	(0.0044)	
T x Quarterly x Post (1st or Second Bill)		-0.0075
		(0.0043)*
Quarterly x Post	-0.0214	-0.0213
	(0.0034)***	(0.0034)***
Quarterly x Post (1st Bill)	0.0204	
	(0.0034)***	
Quarterly x Post (2nd Bill)	0.0244	
	(0.0031)***	
Quarterly x Post (1st or 2nd Bill)		0.0223
		(0.0031)***
T x Bill Number Dummies	Yes	Yes
Bill Number Dummies	Yes	Yes
House Fixed Effects	Yes	Yes
Observations (thousands)	1,456	1,456
R^2	0.0449	0.0449
F Statistic	3705	3877

*, **, ***: Different from zero with 90%, 95%, and 99% confidence, respectively.

Dependent variable is the household's average daily electricity consumption (kilowatt-hours), normalized by average control group consumption in the Post period.

7.5 Heterogeneous Treatment Effects

	I	II	III
T x Post	-0.0186	-0.0189	-0.0189
	(0.0018)***	(0.0018)***	(0.0018)***
T x Quarterly x Post	0.0082	0.0083	0.0084
	(0.0028)***	(0.0028)***	(0.0028)***
Post	-0.0456	-0.0465	-0.0471
	(0.0060)***	(0.0060)***	(0.0060)***
Quarterly x Post	-0.0055	-0.0055	-0.0056
	(0.0020)***	(0.0020)***	(0.0020)***
T x Post x Baseline Usage	-0.0191	-0.0189	-0.0150
	(0.0037)***	(0.0035)***	(0.0033)***
T x Post x Consumer Age	-0.0017		
	(0.0015)		
T x Post x Gas Heat	-0.0073	-0.0071	
	(0.0024)***	(0.0024)***	
T x Post x Household Size	0.0009		
	(0.0016)		
T x Post x House Age	0.0008		
	(0.0015)		
T x Post x log(House Value)	-0.0014		
	(0.0015)		
T x Post x log(Income)	-0.0010		
	(0.0018)		
T x Post x 1(Rent)	-0.0014		
	(0.0013)		
T x Post x Single Family	-0.0004		
	(0.0012)		
T x Post x Square Footage	0.0058	0.0045	
	(0.0018)***	(0.0016)***	
Month x Year Dummies	Yes	Yes	Yes
House Fixed Effects	Yes	Yes	Yes
Post x (X Variable) Controls	Yes	Yes	Yes
Observations (thousands)	1,456	1,456	1,456
R^2	0.0100	0.0100	0.0100
F Statistic	4828	7266	8414

*, **, ***: Different from zero with 90%, 95%, and 99% confidence, respectively.

Dependent variable is the household's average daily electricity consumption (kilowatt-hours), normalized by average control group consumption in the Post period.

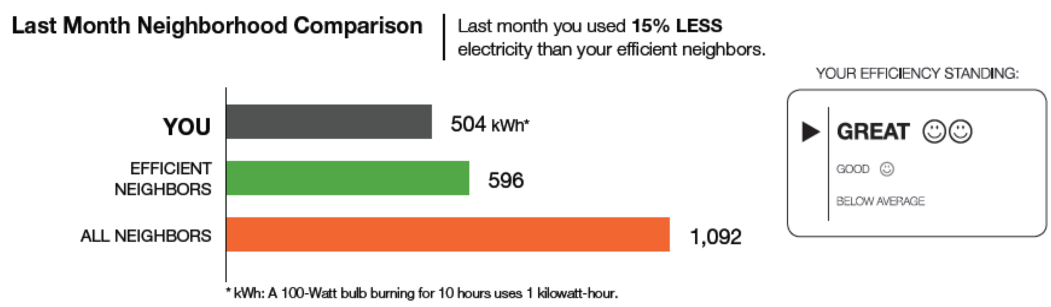
7.6 Profiling and Cost Effectiveness

Assignment Mechanism	*All*	*I*	*II*	*III*
ATE	-0.0191	-0.0343	-0.0346	-0.0347
	(0.0053)	(0.0035)	(0.0035)	(0.0036)
Electricity Bill Savings (dollars/household-year)	20.34	36.56	36.85	36.99
	(5.60)	(3.75)	(3.77)	(3.83)
Cost Effectiveness (cents/kwh saved)	5.31	2.95	2.93	2.92
	(1.46)	(0.30)	(0.30)	(0.30)
Percent Same as II	0.497	0.723		
Percent Same as III	0.498	0.677	0.684	

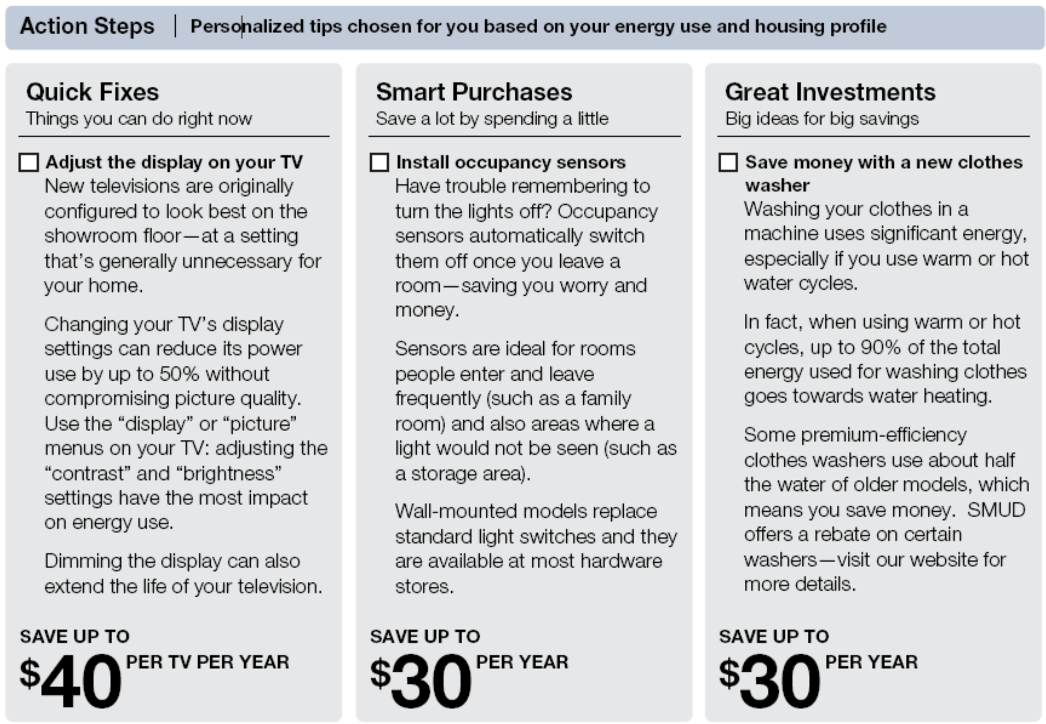
ATE is in percent of Control group usage in the post-treatment period.

8 Figures

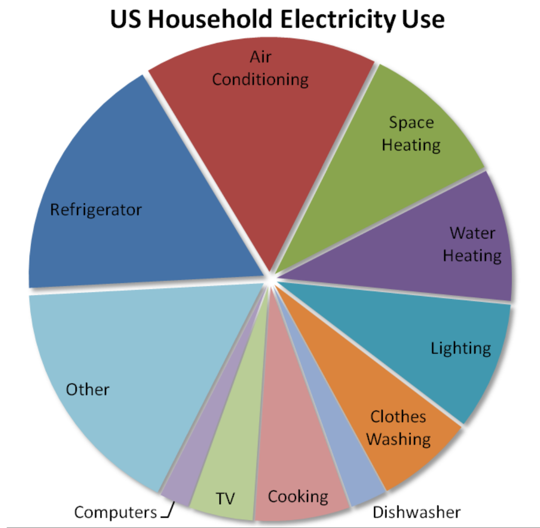
8.1 Home Energy Reports: Social Comparison Module



8.2 Home Energy Reports: Action Steps Module

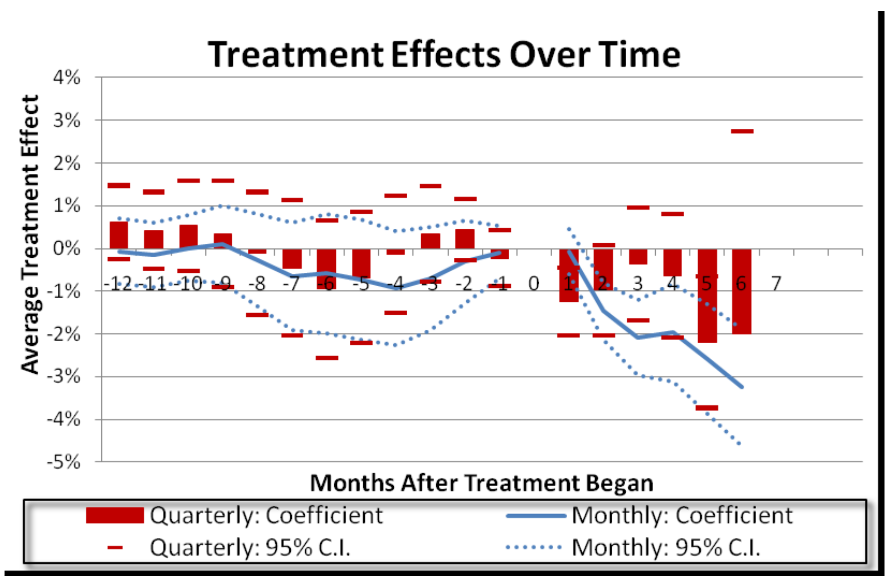


8.3 US Household Electricity Use

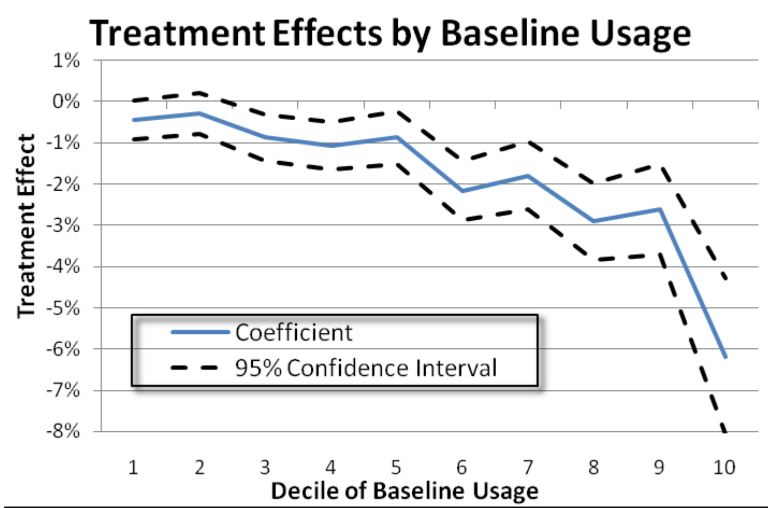


Source: US national average from 2001 Residential Energy Consumption Survey (US Energy Information Administration 2001).

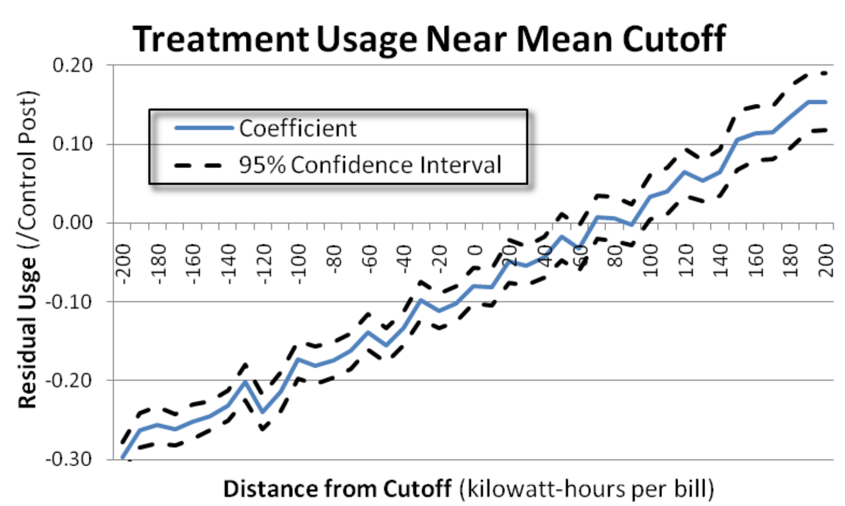
8.4 Treatment Effects Over Time



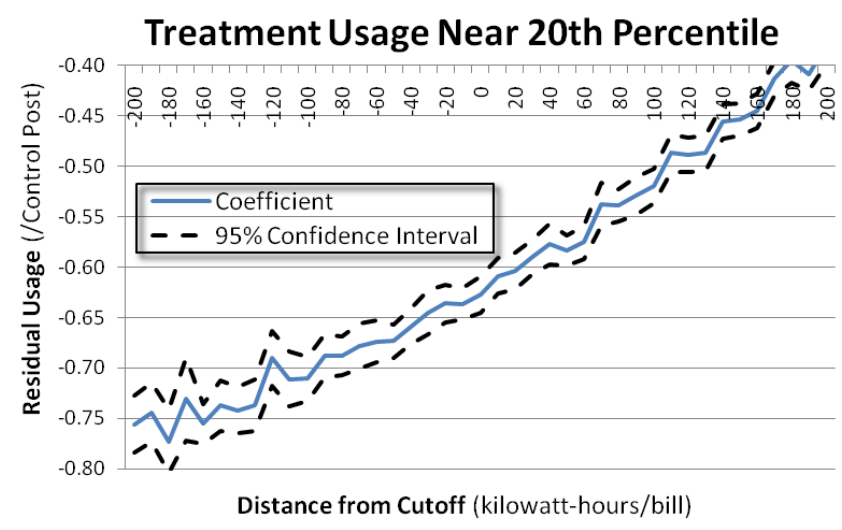
8.5 Treatment Effects by Decile of Baseline Usage



8.6 Treatment Group Near Mean Comparison Cutoff



8.7 Treatment Group Near 20th Percentile Cutoff



8.8 Gains from Profiling