CEEPR Center for Energy and Environmental Policy Research

Social Norms and Energy Conservation

by
Hunt Allcott

09-014 October 2009

A Joint Center of the Department of Economics,
MIT Energy Initiative, and Sloan School of Management

Social Norms and Energy Conservation

Hunt Allcott
MIT and NYU

October 10, 2009

Abstract

This paper evaluates a large-scale pilot program run by a company called OPOWER, previously known as Positive Energy, to mail home energy use reports to residential utility consumers. The reports include energy conservation information as well as social comparisons between a household's energy use and that of its neighbors. Using data from a randomized natural field experiment at 80,000 treatment and control households in Minnesota, I estimate that the monthly program reduces energy consumption by 1.9 to 2.0 percent relative to baseline. In a treatment arm receiving reports each quarter, the effects decay over the months between letters and increase again upon receipt of the next letter. I provide evidence to suggest that at least some of this effect is because consumers' attention is malleable and non-durable. I show that "profiling," or using a statistical decision rule to target the program at households whose observable characteristics predict larger treatment effects, could substantially improve cost effectiveness in future programs. The results provide additional evidence that non-price "nudges" can substantially affect consumer behavior.

JEL Codes: C44, D03, L94, Q41.

Keywords: Social norms, attention, energy demand, randomized field experiments, statistical treatment rules, profiling.

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 Harvard, Stanford, and the Environmental Defense Fund for helpful conversations related to this project. This is a revision of a June 2009 working paper (Allcott 2009a).

1 Introduction

Climate change has emerged as one of the most pressing issues of the early 21st century, and many consider energy efficiency to be a promising approach to reducing energy demand and thus abating greenhouse gas emissions. Historically, economists and policymakers have had the intuition that relative prices are the primary driver of adoption of energy efficient technologies and behaviors. As a result, subsidies for energy efficient capital goods draw the vast majority of federal and state energy efficiency funding (Gillingham, *et al*, 2006).

Perhaps spurred by an increasing general interest in behavioral economics, there has been a recent surge in the appeal of non-price interventions that "nudge" consumers to conserve energy. Non-price interventions are typically inexpensive to implement relative to subsidies, and as demonstrated by Bertrand, *et al*, (2010) in the context of consumer finance, carefully-crafted psychological cues can have effects on consumer behavior that are comparable to large changes in relative prices[1]. A principal challenge, however, is to craft interventions that are scalable and have large effects in a representative population.

This paper econometrically evaluates a large-scale energy conservation program run by a company called OPOWER, previously known as Positive Energy, for an electric utility in Minnesota. The OPOWER program is implemented as a randomized controlled natual field experiment, allowing an unbiased estimate of average treatment effects in the eligible population. The treatment involves mailing to residential consumers Home Energy Reports with two principal features. The first feature is an Action Steps Module that provides information, specifically targeted to each household, on strategies 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 size.

This social comparison feature was motivated by academic work showing that normative comparisons can significantly affect individual behavior. Empirical work by psychologists and political scientists has shown that information on social norms can induce people to conserve energy (Goldstein, Cialdini, and Griskevicius 2008), vote (Gerber and Rogers 2009), or stop littering (Cialdini, Reno, and Kallgren 1990). This social influence literature has developed alongside work by economists on social learning[2] and conditional cooperation in the private provision of public goods[3].

The effects of Positive Energy's program are of practical interest for two reasons. First, independent estimates of the energy use reductions caused by the Home Energy Reports are important

[1] 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.

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

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

per se. The program has been covered in the New York Times (Kaufman 2009), the Atlantic (Tsui 2009), National Public Radio, and other popular media outlets. It has been introduced at 20 utilities, including six of the largest ten in the United States, and by the end of 2009 nearly two million households will be involved nationwide. 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.

Second, the success or failure of the OPOWER 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 controlled trials[4]. If the program has economically significant effects, it would be a remarkable illustration of the effectiveness of non-price interventions compared to the existing large rebates for energy efficient durables.

The point estimates of the Average Treatment Effect (ATE) of OPOWER's monthly Home Energy Reports for the households served by this Minnesota utility range from a 1.9 to a 2.0 percent reduction in electricity consumption relative to the control group. The 95 percent confidence intervals in the four primary specifications span a range from 1.52 to 2.31 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 appear to vary depending on the envelopes in which they are delivered, the time elapsed since the beginning of the treatment, and the season.

The point estimate of the cost effectiveness of the existing Minnesota program is 5.25 cents of program cost per kilowatt-hour saved. This number, which could differ substantially by geographic region and by the scale of the program[5], compares favorably to the price-based approaches of traditional energy efficiency programs. A principal general-interest result of this analysis is the demonstration that a low-cost, non-price intervention - simply sending a letter - can significantly affect consumer behavior. This adds to the growing empirical literature on the power of behavioral interventions, or "nudges," in a variety of domains, including Bertrand, *et al*, (2010), Benartzi and Thaler (2004), Ashraf, Karlan, and Yin (2006), and others.

The experiment was designed to test the efficacy of the overall treatment, not to provide evidence on the mechanisms underlying the effects. The existing data do provide suggestive evidence, however, of several behavioral issues at play. Program households were randomly assigned to groups that would receive Reports every month versus every quarter. Interestingly, the Quarterly group treatment effects decay in the months between receiving Reports, then increase again with the receipt of the next Report, then decay again. The decay is economically significant, as it constitutes almost half of the treatment effect. Survey evidence indicates that an important result of the Reports is to increase day-to-day energy conservation behaviors, such as turning off lights and unplugging appliances, that households *already knew* could save energy. This suggests that the decay in the Quarterly group is the result of a cycle in which receiving a letter reminds or motivates the

[4] OPOWER'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 (e.g. Nadel and Keating (1991)) and are aware of the conceptual benefits of randomized controlled trials, there is no consensus on how the effects of energy efficiency programs should be measured.

[5] Allcott and Mullainathan (2009a, 2009b) calculate a cost effectiveness of 2.9 cents/kilowatt-hour for a nationwide OPOWER program using a simple profiling rule.

household to change behaviors, the attention fades over time, and it is re-engaged upon receiving the next quarter's Report.

The attention channel would be consistent with part of the "Focus Theory of Normative Conduct" in psychology, which originally motivated OPOWER's social comparisons (Cialdini, Reno, and Kallgren 1990). One element of this theory applies bounded attention to social norms, arguing that social norms affect behavior only when at top of mind and suggesting that this attention can dissipate over time. This effect would also be similar to the "Two Steps Forward, One Step Back" dynamic in consumer credit, where credit card fees remind consumers to avoid triggering fees in the future, but that reminder effect decays over months (Agarwal, Driscoll, Gabaix, and Laibson 2006). More generally, the suggestion of bounded and malleable attention is consistent with a growing theoretical and empirical literature in economics, including Chetty, Looney, and Kroft (2009), Finkelstein (2009), Gabaix and Laibson (2006), and others.

Econometric results also show that households that were high energy consumers before the treatment conserve substantially more than households whose baseline consumption was low. This is consistent with models in which low-consumption households have higher marginal costs of energy conservation or are already better informed. Alternatively, this finding is also 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).

This finding of heterogeneous treatment effects also implies that "profiling," or targeting future treatment toward units with highest Conditional Average Treatment Effects, could raise the Average Treatment Effect on the Treated and thus improve the program's cost effectiveness. The final section of the paper builds on this insight to develop a statistical treatment rule under which a decisionmaker who wishes to treat some share of the population in a future program allocates treatment to maximize the expected treatment effect. The routine draws on a recently-growing literature[6] on using the results of randomized experiments to develop statistical decision rules for future treatments. Dehejia (2005), for example, shows that a program that would not be implemented based on examining the ATE alone would increase welfare if implemented in subgroups that have higher Conditional Average Treatment Effects. I show that if the OPOWER program were to be administered to half of the eligible population, profiling would improve its cost effectiveness by a factor of two relative to random assignment. This demonstration of the potential usefulness of profiling is an additional general-interest economic result of the present analysis.

This paper proceeds by providing background on OPOWER's pilot experiment in Minnesota. Section 3 details the econometric strategy, and Section 4 presents results. Section 5 details the statistical treatment rule and gains from profiling, and Section 6 concludes.

2 Experiment Overview

2.1 Motivating Literature

Social scientists have long been aware that social norms can affect individual behavior. 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

[6] This literature includes Berger, Black, and Smith (2000), Dehejia (2005), Graham, Imbens, and Ridder (2009), Hirano and Porter (2006), Imai and Strauss (2009), and Manski (2004, 2009)

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 frugal 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), Axelrod (1984), 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 OPOWER'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.

OPOWER 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 OPOWER pilot experiments is that they take an existing body of scientific knowledge and test whether they are effective at scale, in a natural field experiment in the general population.

2.2 Experimental Design

As of October 2009, OPOWER is operating 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[7]. Minnosota's New Generation Energy Act of 2007 requires that utilities run conservation programs that reduce energy demand by 1.5 percent each year. Although in many states, utilities have little financial incentive to reduce their own sales, such Energy Efficiency Resource Standards have been the primary reason why utilities have contracted with OPOWER.

Connexus Energy has approximately 96,000 households in its service territory. 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 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 9.1 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 9.2, 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.

[7] 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).

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 OPOWER, 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[8].

2.3 Data and Baseline Characteristics

I observe the 1,540,403 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 OPOWER'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 comfortable temperature[9].

Table 8.1 displays the baseline observable characteristics for the Treatment and Control group. Baseline Usage, which is the average across all meter reads in 2008, is 27.7 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 9.3, 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,

[8] 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.

[9] 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.

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[10], 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.

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:

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

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

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.

I define the treatment effect in this way because it is a useful estimand from a policy perspective. OPOWER, 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[11].

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

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

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.

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

3.2 Decay of Effects in Quarterly Group

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 in the Monthly group, which could be driven by seasonal weather changes. I estimate the following equation:

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

The variable B_{12it} is an indicator for whether bill t is the first or second bill after receiving a Report. The first term 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 (or less positive) than τ_Q. The second term 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 8.2 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 first specification is the simple unconditional difference-in-differences estimator. The latter four primary specifications include different configurations of fixed effects, month-by-year dummies, and weather controls; specification III is the exact specification detailed in the Empirical Strategy section. The point estimates of ATEs center around negative 1.9-2.0 percent and 1.5 percent for the Monthly and Quarterly treatments, respectively, and are not statistically different between the five specifications.

Table 8.3 displays two "alternative approaches" to evaluating these pilot programs. The first two columns reflect the approach that would need to be taken if the program did not have a randomized experimental control group. These specifications use only the Treatment group data and implement 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 fourth-order polynomials in Heating and Cooling Degree-Days and 12 month-of-year dummies in the second column, however, the estimated treatment effect is -2.53 percent. This is comparable to but still statistically different from the experimentally-estimated effect with more than 95 percent confidence.

Perhaps the clearest way to emphasize the possible bias from time-varying unobservables is to replicate this regression with the Control group only. In this placebo regression, we expect to see no effects of the treatment. The third column of table 8.3 shows that we estimate a statistically-significant treatment effect on the control group of -0.57 percent. These results underscore the importance of randomized experimentation when results must be measured with precision. Related to the discussion of Lalonde (1986) and Dehejia and Wahba (1999), however, the approximate similarity of the experimental and non-experimental results does suggest that observable time and weather covariates are useful in this setting in controlling for effects on the outcome variable that may be correlated with treatment.

The fourth specification in Table 8.3 presents a cautionary tale about using the logarithmic function. It replicates Specification III from Table 8.2, 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 log specification are 36 and 23 percent lower, respectively, than in the comparable specification in levels. This difference is both economically 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. The percent changes reported in Table 8.2 are that quantity, normalized by Control group consumption in the post-treatment period to generate a percent change. The log specification is fundamentally different: it is the average of the percent reductions across households, instead of the percent of baseline that the average reduction represents. 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, understates the ATE (in absolute value) relative to taking the average change in level across households and then normalizing into a percent.

4.2 Decay of Effects in Quarterly Group

Figure 9.4 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 0, is the meter reading t_{i0} upon which the first reports were based, which was in January or the first few days of February. 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 August, perhaps both because the response to the Reports may grow in the early phases of the program and because the effects may be 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. The point estimate then decreases again in absolute value in month 7.

Table 8.4 presents the formal econometric test, described in the Empirical Strategy section, of whether the effects are stronger for the first two meter reads after receiving a Report (2, 3, 5, and 6) relative to the third (meter reads 4 and 7). 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). Specification II is identical to Specification I except that it combines the dummy variables for the first and second bills.

In this latter specification, the effect of a Quarterly Report relative to the Monthly Report for the first or second bill after the report arrives is -0.15%. The baseline Quarterly variable, which captures the relative effect in the third bill of the quarter, is 0.56%. This is a difference of 0.71%, which is economically substantial: it is nearly half of the Quarterly group's treatment effect. The results for Specification I similarly show that the ATEs for the first and second bills after receiving a Report are stronger than the effect in the third bill of the quarter; they are different with better than 95 percent confidence in a 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. OPOWER 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 capital stock or usage changes that 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 one could model that 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 due to bounded attention.

OPOWER has collected surveys of the treatment group in one of their pilot programs that provide some further evidence on this issue. Treatment group households were asked to self report what they had changed as a result of receiving the Home Energy Reports. Some of the reported effects were indeed seasonal changes to household capital stock, including weather-stripping windows, improving insulation, or servicing the air conditioner. Many of the most frequently reported changes, however, were day-to-day usage behaviors: turning off lights, unplugging electronics, adjusting thermostats, and closing window blinds. Importantly, these behaviors are activities that most consumers likely *already knew* could save them energy[12].

With these data, it is not possible to definitively parse out the information and attention explanations. If day-to-day behavior changes constitute a substantial portion of the treatment effects, however, it is more likely 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 decays over time.

4.3 Effects as a Function of Baseline Usage

Figure 9.5 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 almost zero for the bottom two deciles of baseline usage to 6.4 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 "boomerang effect" model, under which previously

[12] OPOWER is gathering further evidence on both information and behaviors using an improved survey of both treatment and control groups, and these results can be formally presented in the next revision of this paper to strengthen this argument.

low-consumption households might not conserve - or might even consume more - after receiving information that they are less consumptive than their peers.

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. In practice, the Regression Discontinuity estimator does not offer enough power to either estimate a statistically significant effect of the categorizations or to compute a "tightly-estimated zero" by rejecting reasonably small hypothesized effects. See Appendix 7.1 for more details.

5 Profiling and Cost Effectiveness

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, the present problem is straightforward, as the objective function will be unambiguous and the decisionmaker can be modeled as risk-neutral. I focus on perhaps the simplest case of profiling: the OPOWER program is to be allocated to a given proportion of the population, and we want to target the program such as obtain the largest expected treatment effect[13] conditional on the information from a previous randomized trial. The decision variable is which subset of its customer population, defined by observable characteristics, will be assigned to treatment.

The constraint of targeting less than the entire population could arise because the electric utility has limited resources to treat its entire customer base. It could also arise if the Energy Efficiency

[13] 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. Given that costs are constant across individuals, minimizing the cost of achieving a given reduction is equivalent to maximizing the expected treatment effect of treating a given number of people. In designing this system, regulators presumably hope that the solutions to the cost minimization and welfare maximization problems are similar. 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 in the simplest model would be zero.

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.

Resource Standard, the state regulation requiring a given amount of energy conservation relative to baseline, could be satisfied by treating only a portion of the population with the Home Energy Reports. As discussed earlier, state-level Energy Efficiency Resource Standards are the primary reason why utilities contract with OPOWER to reduce the demand for their product.

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 a future OPOWER 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 decisionmaker has information Θ from the existing randomized experiment in a representative sample of larger population $\mathcal{P}$. Recall that, as defined earlier in the paper, the heterogeneous Conditional Average Treatment Effect $\tau(X_i)$ is typically less than zero, meaning that we wish to assign treatment to units with expected treatment effects that are more negative.

The utility's objective is to maximize the expected (negative) treatment effect conditional on the information from the past experiment, given that they are to treat H households from the population:

$$\max_{\delta \in \Delta} E\left[\sum_{i \in \mathcal{P}} -\tau(X_i) \cdot \delta_i | \Theta\right] \quad s.t. \quad \sum_{i \in \mathcal{P}} \delta_i = H \tag{4}$$

Imagine ordering the households in the population by expected treatment effect conditional on their X_i. Denote by R^* the Hth-strongest expected treatment effect. Given the assumption of no spillovers between treatment units, the optimal statistical decision rule collapses to assigning OPOWER treatment to the H households with expected treatment effects more negative than R^*:

$$\delta_i^* = 1\left(E\left[-\tau(X_i)|\Theta\right] \geq -R^*\right) = 1\left(-\widehat{\tau(X_i)} \geq -R^*\right) \tag{5}$$

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 use the frequentist approach more familiar to program evaluation in economics. As suggested in the first line of the above equation, I use the heterogeneous Conditional Average Treatment Effects estimated in OLS, $\widehat{\tau_i(X_i)}$ to construct $E\left[\tau(X_i)|\Theta\right]$. A remaining challenge is to determine which conditioning variables will be used to estimate the heterogeneous treatment effects and to assign future treatment. Including a large set of conditioning variables 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.

Appendix 7.2 details a procedure, based on the work of Imai and Strauss (2009), to determine the optimal set of conditioning variables Z^*. I first define the set of possible conditioning variables $\mathcal{Z}$, which are the observed demographics X plus all interactions and squares of the continuous demographic variables. I then place these potential conditioning variables in order of "prescriptive-ness," which refers to the amount by which conditioning on each individual increases the expected

ATE of a future program. Third, I use cross-validation to generate out-of-sample predictions of the expected ATE while adding progressively less-prescriptive variables to the set of conditioning variables. Fourth, I select the set of conditioning variables that has the largest expected ATE.

This procedure indicates that the optimal set of conditioning variables in this application Z^* actually has only one element: the household's pre-program baseline electricity usage. Conditioning on additional variables causes the model to overfit, reducing the future expected ATE.

5.1 Profiling and Cost Effectiveness: Results

Table 8.5 presents the results of heterogeneous treatment effects regressions for the entire sample, as in Step 2a. 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 1.73 to 2.12 percentage points. The first specification interacts the treatment indicator with the optimal Z^*, the baseline consumption variable.

The second and third regressions represent two alternative "rules of thumb" or selecting prescriptive variables. The second column interacts the treatment indicator variable with all ten available demographic characteristics. Conditional on the other interactions, 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. The third column interacts the treatment indicator only with the three characteristics on which the treatment effect varies with 95 percent confidence in the first specification. The final column is simply the homogeneous treatment effect estimation, where the treatment effect represents the combination of Monthly and Quarterly treatment arms.

Table 8.6 presents the expected ATEs $\widetilde{\tau}(\delta_Z(X))$ when treatment is assigned to one half of the population based on three sets of conditioning variables Z: the optimal set Z^* and two "rule of thumb" conditioning procedures. These ATEs, which pool the effects from the Quarterly and Monthly groups and are computed via the cross validation procedure discussed in Appendix 7.2, are compared to assigning to treatment the entire population or a randomly selected sample thereof. The table shows that profiling doubles the ATE for the treated half of the population. This is remarkable: energy efficiency programs, as well as job training programs, health interventions, and any number of other programs, devote continual effort to incrementally improve their efficacy. These results show that profiling can immediately double the expected average impact.

Interestingly, the expected treatment effects on the treated populations are very similar across the three profiling procedures. In this particular application, the "rules of thumb" for determining covariates on which to condition assignment are nearly as useful as the optimal rule.

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[14]. The average household saves $18 per year on electricity bills as a result of the changes caused by being sent monthly Home Energy Reports.

Using the ATE for the program to date and an estimate of cost per letter sent, I find that the cost effectiveness of the existing Connexus program is 5.25 cents per kilowatt-hour saved. This compares reasonably well to recent point estimates of the average cost of other utility energy efficiency programs, which range from 4.7 to 13.3 cents per kilowatt-hour in a study by Auffhammer,

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

Blumstein, and Fowlie (2008)[15] and from 1.6 to 3.3 cents per kilowatt-hour in a study by Friedrich, *et al*, (2009)[16]. As shown in Table 8.6, using profiling to target treatment at the most responsive half of the population would cut the cost per kilowatt-hour in half, to 2.68 cents-kilowatt-hour[17].

The reader should bear in mind that the ATE, and therefore the cost effectiveness, could vary substantially across utilities and geographic areas, by electricity demand, as determined by season and the mildness of the weather, and as a function of the duration the program has been in effect. For a nationwide rollout of the Home Energy Reports, Allcott and Mullainathan (2009a, 2009b) calculate a somewhat better cost effectiveness, given that program costs would be lower at scale. Some utilities also report that the program affects customer satisfaction, an additional effect that this calculation does not consider. Finally, it is important to re-emphasize that we do not observe how the program changes households' choice of energy services such as comfort from heating and cooling, nor do we observe how their expenditures change for energy-using goods. As a result, we can say little about the welfare effects of the intervention.

6 Conclusion

This paper evaluates the effects of the OPOWER Home Energy Reports, which give households 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 between 1.9 and 2.0 percent below baseline. In the group receiving quarterly Reports, his effect decays somewhat in the months between Reports, either because the information decays seasonally or because the attentional 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.

Although the OPOWER 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. If a utility were limited to

[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 Friedrich, *et al*, (2009) analysis is based on electric utilities' estimates of cost effectiveness, which are typically based on the "deemed savings" measurement approach. Many analysts believe that these estimates could be biased toward zero.

[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 some 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.

implementing the OPOWER program for half of its residential population, I estimate that profiling could make the program twice as cost effective.

Second, this analysis 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 sending a letter - a treatment that has no effect on relative prices and may have limited effects on information sets - can cause measurable changes in demand is remarkable. From the utility's perspective, the cost effectiveness of the OPOWER program compares favorably to the traditional large subsidies for energy efficient durable goods. Perhaps the most important conclusion, therefore, is that some combination of information, attention, 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). "Rethinking Real Time Electricity Pricing." MIT Center for Energy and Environmental Policy Working Paper 2009-015 (October).

[7] Allcott, Hunt (2009a). "Attention, Social Norms, and Energy Conservation." Working Paper, Massachusetts Institute of Technology (June).

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

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

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] Axelrod, Robert (1984). The Evolution of Cooperation. New York: Basic Books.

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

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

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

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

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

[20] 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.

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

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

[23] 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.

[24] 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.

[25] 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).

[26] 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.

[27] 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.

[28] 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.

[29] Chetty, Raj, Adam Looney, and Kory Kroft (2009). "Salience and Taxation: Theory and Evidence." *American Economic Review*, Vol. 99, No. 4 (September), pages 1145-1177.

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

[31] 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.

[32] 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.

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

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

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

[36] 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.

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

[38] Dehejia, Rajeev, and Sadek Wahba (1999). "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs." *Journal of the American Statisitcal Association*, Vol. 94, No. 448 (December), pages 1053-1062.

[39] 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.

[40] 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.

[41] 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.

[42] Finkelstein, Amy (2009). "E-ZTAX: Tax Salience and Tax Rates." *Quarterly Journal of Economics*, Vol. 124, No. 3 (August), pages 969-1010.

[43] 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.

[44] 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.

[45] 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.

[46] Friedrich, Katherine, Maggie Eldridge, Dan York, Patti Witte, and Marty Kushler (2009). "Saving Energy Cost-Effectively: A National Review of the Cost of Energy Saved through Utility-Sector Energy Efficiency Programs." ACEEE Report No. U092 (September).

[47] Gabaix, Xavier, David Laibson (2006). "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets." *Quarterly Journal of Economics*, Vol. 121, No. 2 (May), pages 505-540.

[48] 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.

[49] 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.

[50] 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.

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

[52] 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.

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

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

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

[56] 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.

[57] 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.

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

[59] 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.

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

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

[62] Lalonde, Robert (1986). "Evaluating the Econometric Evaluations of Training Programs." *American Economic Review*, Vol. 76, No. 4 (September), pages 604-620.

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

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

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

[66] 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.

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

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

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

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

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

[72] 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.

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

[74] 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.

[75] 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.

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

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

[78] 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.

[79] 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.

[80] 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.

[81] 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.

[82] 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.

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

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

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

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

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

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

[89] Wald, Abraham (1950). <u>Statistical Decision Functions</u>. New York, NY: John Wiley & Sons.

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

[91] 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 Appendices

7.1 Regression Discontinuity

In this Appendix, which need not be included for publication, I briefly discuss a Regression Discontinuity specification that could in principle be used to estimate the relative effects of the different normative categorizations Social Comparison Module in the Home Energy Reports. 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 9.8 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. The line is upward sloping, as households that consume less compared to their peers on a given bill also tend to have lower residual usage'on future bills. Figure 9.9 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 of Control group post-treatment consumption, which is 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 OPOWER'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.

7.2 Variable Selection

This Appendix details the procedure for selecting a vector of observables and interactions to condition on when estimating $\widehat{\tau_i(X_i)}$ in the profiling procedure. The discussion of this problem and my approach to addressing it draw heavily on previous work by Imai and Strauss (2009) and, somewhat less directly, on Gunter, Zhu, and Murphy (2007).

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 by definition correlated with the outcome and therefore may be useful for reducing residual variance, i.e. improving the efficiency of the estimator. However, if a predictive variable has the same correlation with the outcome in treatment and control, meaning that it is not correlated with the treatment effect, this variable would not be useful for generating a statistical decision rule.

Most fundamentally, a powerful prescriptive variable is one that increases the efficacy of the targeted program in a future implementation when profiling is conditioned on that variable. Intuitively, one approach to choosing the set of prescriptive variables is to try conditioning on all possible covariates and subsets thereof, generating assignment rules based on each potential set of covariates, then running a future program based on that assignment rule and measuring the effects. Because these theoretical future programs differ only in the rules used to assign treatment, the future program with the highest ATE is the one with the best assignment rule.

The actual approach is designed to mimic that process, while addressing two concerns. First, we of course cannot try many different future implementations. We can, however, mimic that process with the data from the existing program: we can partition the data, use the "first" part to generate the treatment rule and the "second" part to test its efficacy. If we then reverse the partition - use the "second" part to generate the treatment rule and the "first part" to test efficacy, we will have then exploited all of the data available while still cross-validating the assignment mechanism outside of the sample that generated it.

A second problem with the intuitive approach is that it would be very computationally intensive to try all possible sets of covariates. Instead, using a procedure developed by Imai and Strauss (2009), we can rank the individual covariates based on their "prescriptiveness." In this context, "prescriptiveness" means how much conditioning and assigning based on that one covariate increases the expected treatment effect relative to random assignment of treatment.

Imagine ordering the variables from most to least prescriptive. Beginning with the most prescriptive variable, one can progressively add the next-most prescriptive variable to the set of conditioning variables. Conditioning assignment on additional covariates should for some time increase the future treatment effect of the future targeted program, but the overfitting effect will eventually dominate and the future treatment effect will decrease. This procedure of sequentially adding variables in order of prescriptiveness ensures that we have the most prescriptive set of variables at the apparent maximum. This ordering procedure simplifies the problem from one of selecting the optimal *set* of covariates to selecting the optimal *number* of covariates. Because of the ordering, the optimal number will also tell us the optimal set (Imai and Straus 2009).

7.2.1 Specifics of the Variable Selection Algorithm

The goal is to select the optimal set of covariates Z^*, drawn from some set of possible covariates $\mathcal{Z}$, such that a statistical treatment rule conditioning on Z^* maximizes the expected treatment effect in a future implementation of the program. The variables Z are functions of household observable characteristics X. Denote by $\delta_Z(X)$ the treatment rule using variables Z, and denote by $\widetilde{\tau}(\delta_Z(X))$ the expected treatment effect of a future program using that rule. Consistent with the maximization problem introduced above, that treatment effect is conditioned on including some set of people such that the minimum treatment effect is above some threshold R^*. For simplicity, in this part of the problem, I assume that $\widetilde{\tau}(\delta_Z(X))$ reflects the treatment effect when one-half of the population is to be treated. The maximization problem is:

$$Z^* = \arg\max_{Z \in \mathcal{Z}} E\left[-\widetilde{\tau}(\delta_Z(X))|\Theta\right] \tag{6}$$

Following Imai and Strauss (2009), I implement a four-step procedure.

Step 1: Define the possible set of prescriptive variables. In principle, the possible set of prescriptive variables $\mathcal{Z}$ could be all levels of interactions of the 10 observable household characteristics, plus polynomial series of the continuous variables. To speed computation, I use only the variables themselves, interactions of each pair, and squares of the continuous variables. This gives a set $\mathcal{Z}$ of $J = 62$ possible covariates, which are indexed $j = 1, ..., J$.

Step 2: Order the set of variables by prescriptiveness. In this step, the set of J possible covariates is ordered by "prescriptiveness": the amount by which conditioning on that one variable Z_j increases the expected future treatment effect $\widetilde{\tau}(\delta_{Z_j}(X))$.

For each of the J possible covariates Z_j, the procedure is:

Step 2a: Use the entire dataset to estimate the average treatment effect, combining Monthly and Quarterly treatment arms, and the amount by which the ATE changes with a change in Z_j:

$$Y_{it} = (\tau_0 + \tau_1 \cdot Z_j) \cdot T_i P_{it} + \beta \cdot P_{it} + \mu_{my} + \upsilon_i + \varepsilon_{it} \tag{7}$$

Step 2b: For both the Treatment and Control groups, fit each individual i's $\widehat{\tau(X_i)} = \widehat{\tau}_0 + \widehat{\tau}_1 \cdot Z_{ji}$ from the above regression.

Step 2c: Treatment rule $\delta_{Z_j}(X)$ assigns treatment to the one-half of households with the most negative fitted treatment effects.

Step 2d: Using the half of the original randomized Treatment and Control groups assigned to treatment by this potential rule, estimate the treatment effect $\widetilde{\tau}(\delta_{Z_j}(X))$.

After computing these expected average treatment effects for each Z_j, these effects are ordered from the largest negative treatment effect to smallest. This gives a new ordering of the set $\mathcal{Z}$ from most to least prescriptive, such that $-\widetilde{\tau}(\delta_{Z_j}(X)) \geq -\widetilde{\tau}(\delta_{Z_{j+1}}(X)), \forall j$. The double solid line in Figure 9.6 illustrates this ordering for $j = 1, ..., 62$. The most prescriptive variable is Baseline Usage, followed by interactions of Baseline usage with the Single Family and Rent indicator variables.

Step 3: Determine the expected ATE for each combination of prescriptive variables. For each combination of prescriptive variables with effectiveness greater than or equal to that of j, cross validation is now used to predict the expected treatment effect of profiling on that combiation in assigning treatment for a future program. First, the sample is randomly divided into $K = 5$ subsets for K-fold cross validation.

For each value of j from 1 to J, which are now ordered from most to least prescriptive, the procedure is:

Step 3a: Define $Z_{1,...,j}$ as the set of variables such that $-\widetilde{\tau}(\delta_{Z_l}(X)) \geq -\widetilde{\tau}(\delta_{Z_j}(X))$. This set contains all the variables, including j, that have at least as prescriptive as j. For each of the K cross-validation sets, estimate the heterogeneous treatment effects equation using the $K-1$ "training sets" excluding k. In set k, which is out of sample, then determine the treatment rule $\delta_{kZ_{1,...,j}}(X)$.

Step 3b: After using cross validation to determine the treatment rule $\delta_{Z_{1,...,j}}(X)$ for all Treatment and Contol observations in the data, then use the original randomization into Treatment and Control to estimate the ATE for those units that would be assigned to treatment, $\widetilde{\tau}(\delta_{Z_{1,...,j}}(X))$.

Step 4: Determine the optimal number of variables. The single solid line appearing on the lower portion of Figure 9.6 illustrates the $\widetilde{\tau}(\delta_{Z_{1,...,j}}(X))$ computed via cross validation when profiling based on an increasing number of covariates Z. The optimal set of covariates Z^* is determined by the point on that line with the largest (negative) treatment effect:

$$j^* = \arg\max_j E[-\widetilde{\tau}(\delta_{Z_{1,...,j}}(X))|\Theta] \qquad (8)$$

In this case, the most negative treatment effect is at the far left of Figure 9.6, meaning that $j^* = 1$. In this application, the optimal statistical treatment rule assigns treatment based only on Baseline Usage.

8 Tables

8.1 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	31.57	31.40	0.166	32.07	30.99	31.59
	(28.03)	(27.96)	(0.200)	(28.06)	(27.68)	(28.24)
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	1659	1.41	1614	1637	1703
	(454)	(449)	(3.22)	(442)	(441)	(464)
F-Test p-Value			0.384			
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	769299	771083		276,922	650,671	609,982

*, **, ***: 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.

8.2 Treatment Effects

	I	II	III	IV	V
T x Post	-0.0152	-0.0192	-0.0191	-0.0197	-0.0189
	(0.0045)***	(0.0019)***	(0.0019)***	(0.0018)***	(0.0019)***
T x Quarterly x Post	-0.0019	0.0045	0.0045	0.0044	0.0043
	(0.0055)	(0.0024)*	(0.0024)*	(0.0022)**	(0.0024)*
Post	-0.0919	-0.0841	-0.0453	-0.0870	-0.0058
	(0.0022)***	(0.0011)***	(0.0058)***	(0.0050)***	(0.0058)
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)	1,540	1,540	1,540	1,540	1,540
R^2	0.0041	0.0041	0.0453	0.0000	0.0485
F Statistic	4159	4284	7053	3391	3018

*, **, ***: 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.

8.3 Alternative Empirical Approaches

	Treatment Only	Treatment Only	Control Only	In Logs
T x Post				-0.0122
				(0.0016)***
T x Quarterly x Post				0.0009
				(0.0021)
Post	-0.1033	-0.0253	-0.0057	-0.0430
	(0.0015)***	(0.0017)***	(0.0016)***	(0.0039)***
Quarterly x Post	0.0045	0.0043	-0.0045	-0.0373
	(0.0024)*	(0.0024)*	(0.0023)**	(0.0009)***
Degree-Day Polynomial		Yes	Yes	
Month Dummies		Yes	Yes	
Month x Year Dummies				Yes
House Fixed Effects	Yes	Yes	Yes	Yes
Observations (thousands)	769	769	771	1540
R^2	0.0047	0.0479	0.0484	0.0553
F Statistic	3675	1249	1252	13362

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

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

8.4 Decay of Quarterly Treatment Effects

	I	II
T x Quarterly x Post	0.0056	0.0056
	(0.0034)	(0.0034)
T x Quarterly x Post (1st Bill)	-0.0009	
	(0.0032)	
T x Quarterly x Post (2nd Bill)	-0.0021	
	(0.0024)	
T x Quarterly x Post (1st or Second Bill)		-0.0015
		(0.0025)
Degree-Day Polynomial		
T x Bill Number Dummies	Yes	Yes
Bill Number Dummies	Yes	Yes
House Fixed Effects	Yes	Yes
Observations (thousands)	1,456	1,540
R^2	0.0431	0.0431
F Statistic	3795	3883

*, **, ***: 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.

8.5 Heterogeneous Treatment Effects

	I	II	III	IV
T x Post	-0.0172	-0.0168	-0.0172	-0.0173
	(0.0015)***	(0.0014)***	(0.0015)***	(0.0016)***
T x Post x Baseline Usage	-0.0173	-0.0212	-0.0209	
	(0.0036)***	(0.0039)***	(0.0038)***	
T x Post x Consumer Age		-0.0018		
		(0.0015)		
T x Post x Gas Heat		-0.0063	-0.0061	
		(0.0024)***	(0.0024)**	
T x Post x Household Size		0.0020		
		(0.0017)		
T x Post x House Age		0.0004		
		(0.0016)		
T x Post x log(House Value)		-0.0012		
		(0.0015)		
T x Post x log(Income)		-0.0009		
		(0.0019)		
T x Post x 1(Rent)		-0.0006		
		(0.0014)		
T x Post x Single Family		-0.0003		
		(0.0013)		
T x Post x Square Footage		0.0053	0.0043	
		(0.0019)***	(0.0017)**	
Post	-0.0470	-0.0438	-0.0447	-0.0453
	(0.0060)	(0.0060)***	(0.0060)***	(0.0058)***
Month x Year Dummies	Yes	Yes	Yes	Yes
House Fixed Effects	Yes	Yes	Yes	Yes
Post x (X Variable) Controls	Yes	Yes	Yes	-
Observations (thousands)	1,540	1,540	1,540	1,540
R^2	0.0040	0.0041	0.0040	0.0453
F Statistic	8506	4769	7216	7374

*, **, ***: 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.

The ATE represents combination of Monthly and Quarterly groups.

Specification I: Optimal set of conditioning variables Z^*: Baseline Usage only.

Specification II: "Rule of thumb" conditioning variables: All observed demographic variables

Specification III: "Rule of thumb" conditioning variables: Only demographic variables statistically significantly associated with the treatment effect.

Specification IV: No conditioning variables.

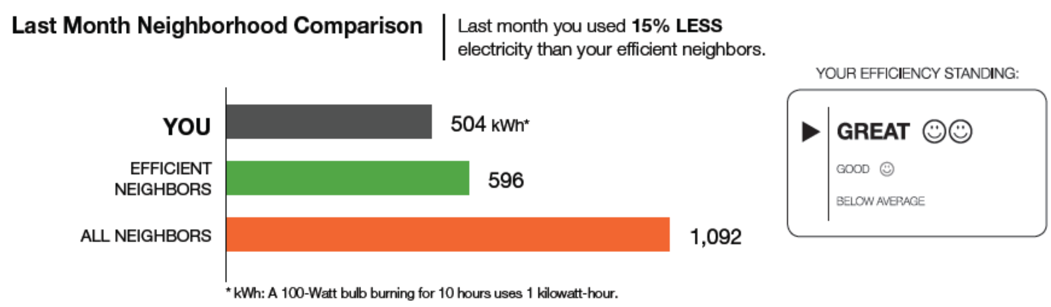
8.6 Profiling and Cost Effectiveness

Assignment Mechanism	*All*	*I*	*II*	*III*
Expected ATE (percent)	-0.0173	-0.0338	-0.0322	-0.0328
	(0.0016)	(0.0033)	(0.0032)	(0.0032)
Electricity Bill Savings (dollars/household-year)	18.42	36.05	34.32	34.89
	(1.74)	(3.49)	(3.44)	(3.45)
Cost Effectiveness (cents/kwh saved)	5.25	2.68	2.82	2.77
	(0.50)	(0.26)	(0.28)	(0.27)
Percent of Households Assigned Same as II	0.497	0.678		
Percent of Households Assigned Same as III	0.497	0.690	0.717	

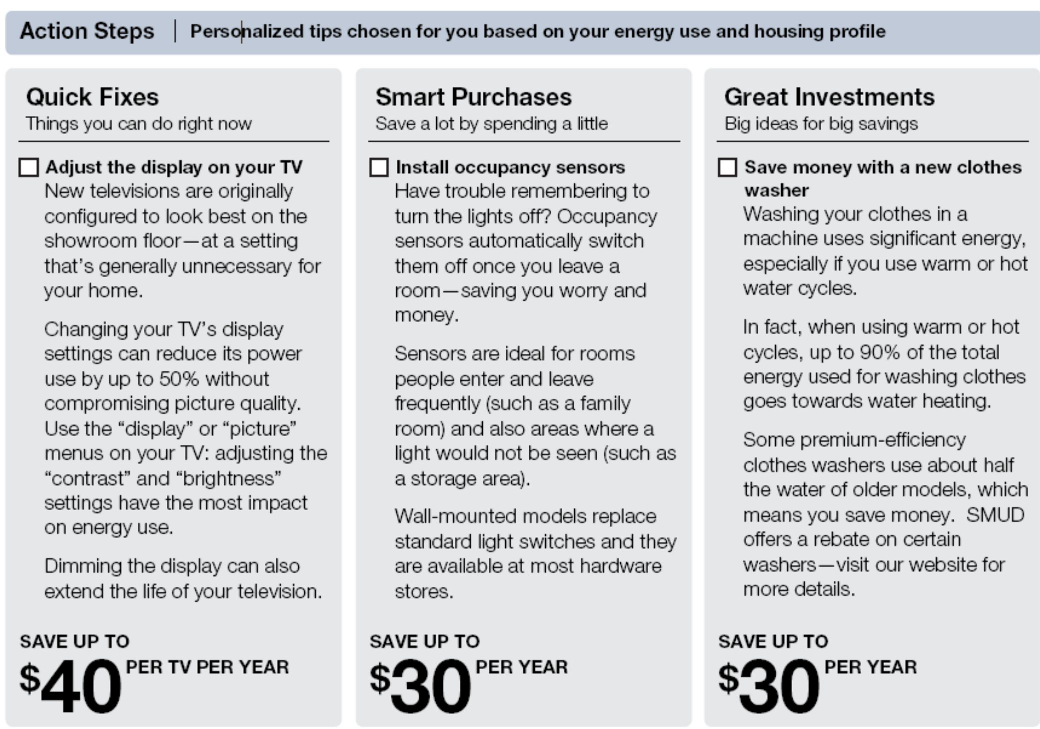
ATE is in percent of Control group usage in the post-treatment period.
The ATE represents combination of Monthly and Quarterly groups.
Assignment Mechanism "All": Assign to treatment all households, or a randomly-selected subset.
Assignment Mechanism I: Optimal set of conditioning variables Z^*: Baseline Usage only.
Assignment Mechanism II: "Rule of thumb" conditioning variables: All observed demographic variables
Assignment Mechanism III: "Rule of thumb" conditioning variables: Only demographic variables statistically significantly associated with the treatment effect.

9 Figures

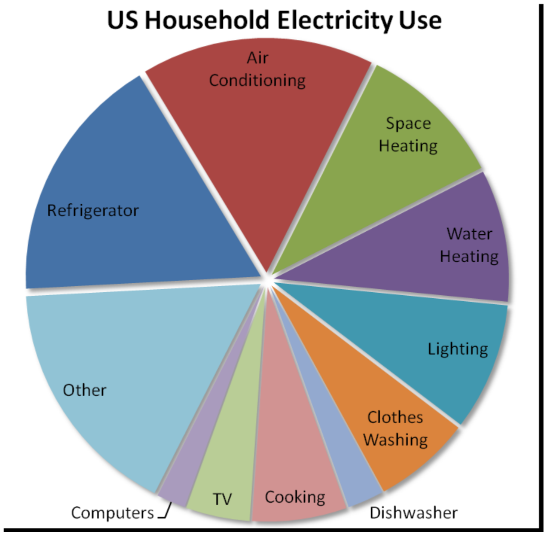
9.1 Home Energy Reports: Social Comparison Module



9.2 Home Energy Reports: Action Steps Module

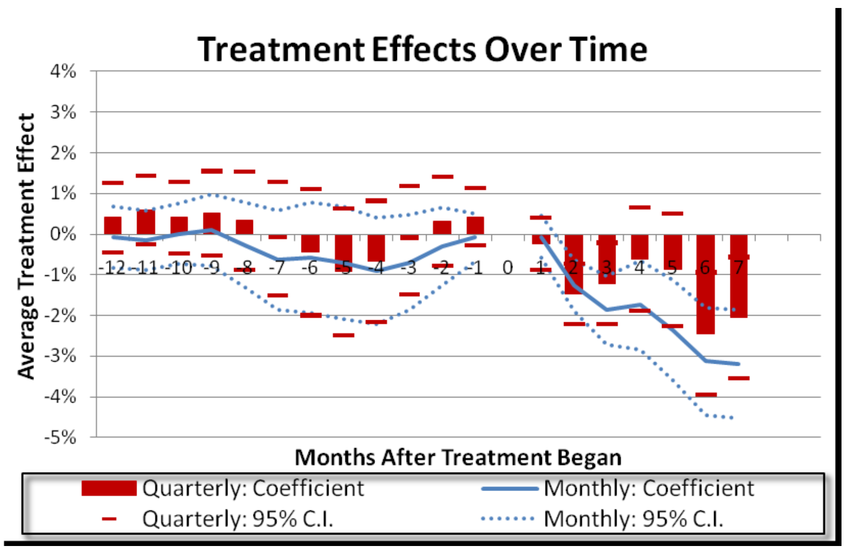


9.3 US Household Electricity Use

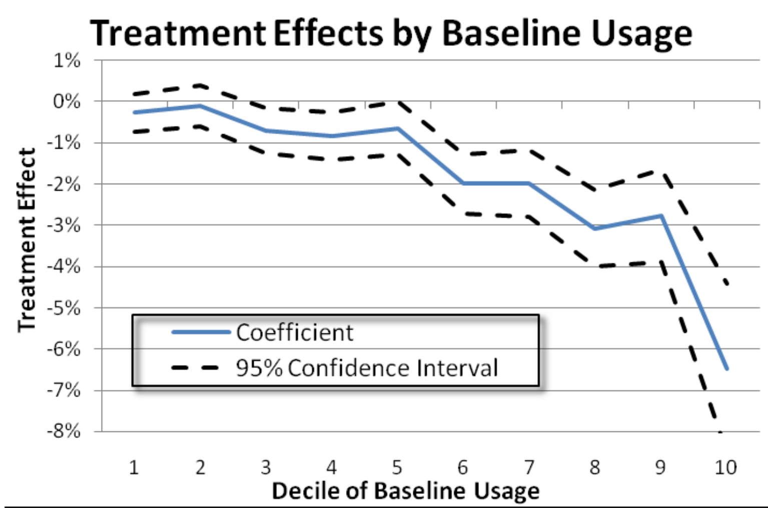


Source: US national average from 2001 Residential Energy Consumption Survey (US Energy Information Administration 2001).

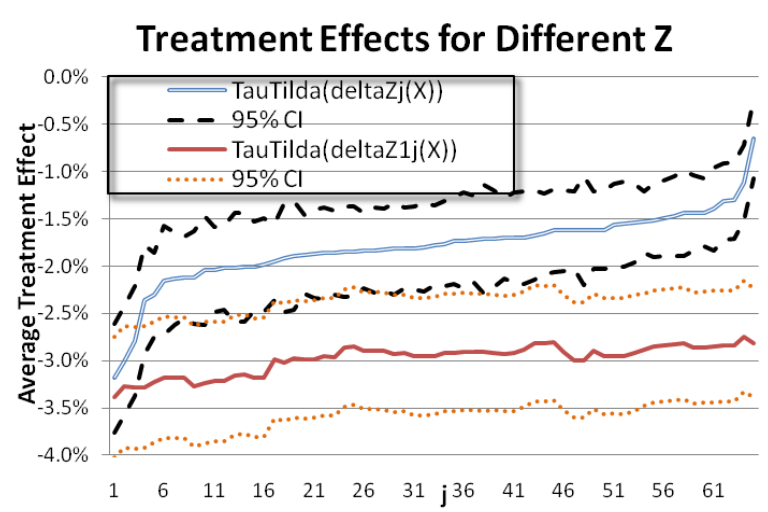
9.4 Treatment Effects Over Time



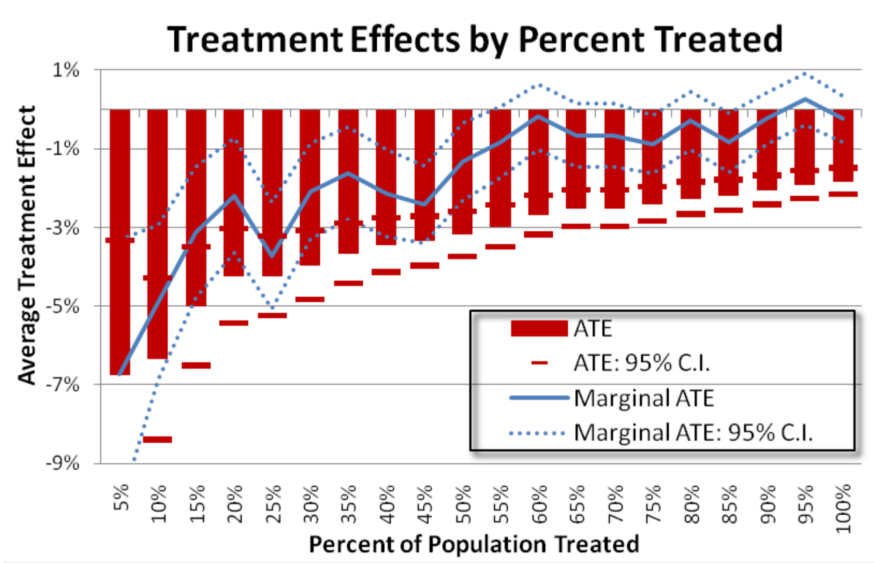
9.5 Treatment Effects by Decile of Baseline Usage



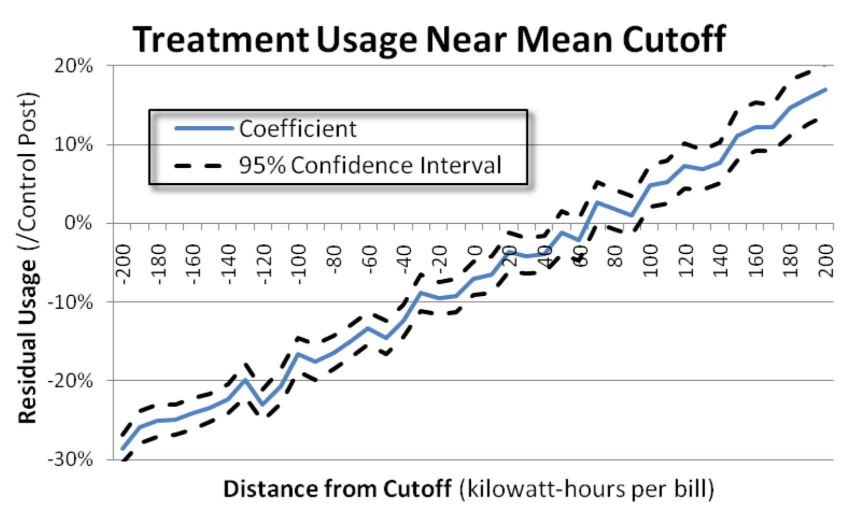
9.6 Profiling: Treatment Effects for Different $\mathbf{Z}_j$



9.7 Gains from Profiling



9.8 Treatment Group Near Mean Comparison Cutoff



9.9 Treatment Group Near 20th Percentile Cutoff