Using Targeting to Optimize Program Design: Evidence from an Energy Conservation Experiment

Todd D. Gerarden and Muxi Yang*

August 2022

Abstract

We investigate the potential for targeted treatment assignment rules to improve the performance of a large-scale behavioral intervention to encourage households to conserve energy. We derive treatment rules based on observable household characteristics that maximize the expected benefits of the intervention. Targeting treatment using transparent and easily implemented rules could yield significant gains; the energy savings from optimal treatment assignments are predicted to be double those achieved by the intervention as implemented. Predicted cost savings from targeting are even larger. Our results underscore the potential for targeted treatment assignment to generate significant benefits in many domains.

Keywords: energy conservation, randomized field experiments, targeted treatment
JEL Codes: Q41, C44, C93

*Gerarden: Cornell University. gerarden@cornell.edu. Yang: Cornell University. my458@cornell.edu.
We thank Hunt Allcott, Abhishek Ananth, Karen Brandon, Toru Kitagawa, Chris Knittel, Shanjun Li, Erica Myers, Zhuan Pei, Brian Prest, Jörg Stoye, Aleksey Tetenov and seminar participants at Cornell, the Triangle Resource and Environmental Economics Seminar, the Columbia-Cornell-MIT-Stanford EEE Seminar, the Berkeley-Harvard-Yale EEE Virtual Seminar, the Eastern Economic Association Annual Conference, the University of Wisconsin-Madison AAE Department Seminar, the AERE Summer Conference, IAEE 2021, the Northeast Workshop on Energy Policy and Environmental Economics, the LSE Workshop in Environmental Economics, the EAERE Annual Conference, and AERE@ASSA 2022 for comments.

1 Introduction

The central goal of program evaluation is to identify and design interventions that maximize welfare criteria. With this in mind, a vast literature of prior work has focused on identifying programs that generate positive net benefits. Recent methodological advances, however, present an opportunity to take this further, by optimizing the design of programs to *maximize* those net benefits. One line of research has focused on the effects of treatment assignment policies – the rules that govern which individuals should or should not be treated. This research has provided a theoretical foundation and a framework for thinking about ways to exploit treatment effect heterogeneity to maximize the benefits of policy interventions, with the potential to generate large welfare gains in many domains.

In this study, we use policy learning methods to examine ways to maximize benefits from programs that aim to encourage household energy conservation. We analyze whether simple and transparent treatment assignment rules could improve upon the outcomes achieved through a large-scale behavioral intervention undertaken by a public utility in the Northeast United States. In the program, the public utility mailed personalized letters to households to give them feedback and social comparisons on energy consumption. The aim of the program was to reduce household energy consumption by addressing imperfect information and behavioral biases. Such "home energy report" programs have been adopted by utilities across the United States, reaching at least 6.2 million households (Allcott and Rogers, 2014). The sheer scale of these programs warrants examining ways to optimize their design; indeed, if such programs were expanded to include all households in the country, the annual costs could exceed $1.2 billion.[1] Furthermore, home energy reports are one of the few residential energy-efficiency programs that have been shown to be cost-effective (see, e.g., Allcott, 2011). Prior research on home retrofits (Fowlie et al., 2018) and appliance replacement programs (Davis et al., 2014) have found that the benefits of these interventions do not exceed their costs on average.[2] Given this, and given the centrality of energy efficiency to climate policy, it is important to design home energy report programs to maximize their net benefits.

[1]The estimated annual cost is calculated by updating numbers from Allcott and Mullainathan (2010): multiplying $7.48 per household by 135 million households in 2019, as reported by the U.S. Energy Information Administration, and then inflating to 2020 dollars.

[2]Programs that do not appear cost-effective on average can also benefit from targeting. As an example, Christensen et al. (2021) find that less than half of homes treated by a home energy retrofit program had positive net private benefits, and that targeting funds to a subset of homes could shift the program from producing net negative to net positive social benefits. In principle, tailoring such interventions to treat only the households for which benefits exceed costs could prove as or more valuable than the targeting of interventions like home energy reports for which benefits already exceed costs on average.

Home energy reports have been widely implemented as randomized controlled trials. The use of such experiments has been integral in establishing the causal short- and long-term impacts of home energy reports on household energy consumption (see, for example, Allcott, 2011; Allcott and Rogers, 2014). However, repeated randomized trials are unlikely to be welfare maximizing. One reason lies in the heterogeneous treatment effects that the literature has uncovered. For example, prior work has found that certain households could *increase* electricity consumption after receiving positive feedback in informational letters (Schultz et al., 2007; Byrne et al., 2018). Due to this potential adverse response, and given that such behavioral programs incur implementation costs, it could be welfare enhancing to utilize information from prior randomized trials to identify and target households that "should" be treated based on observable characteristics. We examine this issue by leveraging data on household-level monthly electricity consumption, treatment status, and five demographics characteristics from a program with over 390,000 participants. We apply new statistical learning methods to these data to identify treatment assignment rules that would maximize the program's net benefits, and we provide an estimate of the size of the potential benefits that would accrue if such measures were put in place.

We motivate our analysis of the potential gains that could stem from targeting by providing evidence of heterogeneous treatment effects. We estimate triple-difference regressions that provide suggestive evidence of how treatment effects may vary across households. In particular, we estimate separate treatment effects for households who are above and below the median in terms of four key observable characteristics: pre-treatment electricity consumption levels, household income, house size, and house age. This reveals certain treatment effect patterns. Households with high pre-treatment consumption reduce electricity usage more than households with low pre-treatment consumption. A similar picture emerges in terms of income; higher-income households respond to the treatment to a greater degree than lower-income households. There is relatively less heterogeneity in treatment effects in terms of house size and age. Households with larger and older homes appear to exhibit slightly larger treatment effects, but the differences in the effects that emerge between these households and those in smaller and newer houses are not statistically significant. We use this suggestive evidence on heterogeneous treatment effects to guide the selection of covariates for our targeting analysis.

In our main analysis, we use empirical welfare maximization (Kitagawa and Tetenov, 2018) to search for simple and transparent treatment rules that maximize the net benefits of the intervention. This method leverages randomization of treatment in the original experiment to estimate the effects of alternative treatment assignment rules. These alternative rules assign households to treatment based on their observable characteristics. To implement this method, we restrict our attention to two household characteristics at a time. We then search over two types of treatment rules. First, we search over quadrant partitions of each two-dimensional characteristic space to identify the quadrant that, when treated, maximizes expected benefits. Second, we use a linear rule with cubic terms to allow for a more flexible partitioning of the characteristic space. In both cases, we search for rules using three separate criteria: energy conservation, private cost savings, and social cost savings. For the cost-savings analyses, we account for both the value of electricity conserved and the cost of administering the program. We do not observe costs incurred by households that receive the letters, though, so our results should not be interpreted as welfare impacts.[3] In each analysis, our approach yields a rule that determines the optimal treatment assignment for each household based on its observable characteristics. The rules are easy to implement, and they can be visualized in two dimensions.

We find large gains in cost-effectiveness from using observable household characteristics to target treatment. When targeting treatment assignment to maximize energy conservation, our estimates suggest that the predicted reduction in electricity consumption could double the consumption reduction that was generated by the program as it was implemented. Using targeting to maximize cost savings has even greater potential to improve program design. We find that alternative treatment assignment rules could achieve 65 to 82 cents in private cost savings per household per month, net of the program's implementation costs. This equates to roughly \$385,000 to \$485,000 in total net cost reductions per year for the sample. By contrast, we estimate that the program as implemented generated net cost reductions of only \$50,000. These results imply that targeting could have increased the net cost savings generated by the program by an order of magnitude.

This is one of the first studies that derives targeted treatment assignment rules for an energy information provision program. Our results confirm the potential of targeting treatment to improve program performance in aggregate. Targeting has potential downsides, though. Since targeting treats different households differently, it may generate inequitable outcomes. Using household income or other demographic characteristics to target treatment may also be politically challenging. Demographic data may not always be available. We address these concerns in two ways. First, we examine the performance of targeting on the basis of pre-treatment electricity consumption alone. We find that treatment rules derived only using data on pre-treatment electricity consumption

[3]Furthermore, because the frequency of our electricity consumption data is monthly, our analysis does account for intra-month variation in the social benefits of electricity conservation.

perform as well as rules that depend on household characteristics such as income. Second, we document how targeted treatment assignment would affect households of different incomes and races. We find that some rules disproportionately treat high-income households, while others treat households across the income distribution. Program administrators could take these distributional impacts into account by using them to select among alternative treatment assignment rules, or by incorporating them directly into the optimization problem used to derive treatment assignment rules.

A distinguishing characteristic of our work is that this is the first paper to conduct inference on the potential gains from targeted assignment of home energy reports. Prior work has presented point estimates of these potential gains, but has not accounted for uncertainty in those estimates. We do, and in doing so we provide the first evidence of statistically significant gains from targeting the assignment of home energy reports. Thus, our approach could be used to propose treatment rules that could be implemented easily, and would improve the cost-effectiveness of energy conservation programs.

Related Literature We build on a large literature that studies the causal impacts of home energy reports on household energy consumption (see, for example, Allcott, 2011, 2015; Ayres et al., 2013; Allcott and Rogers, 2014; Brandon et al., 2017, 2019). In particular, our analysis of targeting is motivated by prior evidence of heterogeneity in treatment effects found by Allcott (2011) and Byrne et al. (2018), among others. These heterogeneous treatment effects can be driven by matters such as heterogeneity in both household preferences and energy market distortions (Costa and Kahn, 2013; Allcott et al., 2015; Myers and Souza, 2020). We contribute to this literature by studying how heterogeneity in how households respond could be leveraged to optimize the design of home energy report programs.

The two most closely related papers from the literature on home energy reports are Allcott and Kessler (2019) and Knittel and Stolper (2019). In both studies, the authors use machine learning methods to estimate household-specific conditional average treatment effects. They then use these treatment effect predictions to quantify the potential impacts of treating only the households who respond most favorably to treatment. Our research question is similar, but our method is distinct. One advantage of our approach is that the treatment rules are simple and transparent. This could facilitate adoption of targeting by utilities and regulators. A second important feature of the method we use is that it integrates the decision problem and statistical inference. This is conceptually appealing, and it allows us to construct confidence intervals for our estimates of the potential gains from targeting. In contrast, prior work only provides point estimates.

Finally, we utilize recently developed methods from the econometrics literature on statistical treatment rules. We rely most heavily on Kitagawa and Tetenov (2018), who introduce the empirical welfare maximization method that we apply to home energy reports. This method is grounded in minimization of the maximum welfare loss from failing to adopt ideal treatment rules, or minimax regret. The use of minimax regret as a decision criterion has its origins in statistical decision theory (Wald, 1950; Savage, 1951), and recent developments provide tractable methods to derive treatment rules based on this criterion (e.g., Manski, 2004; Hirano and Porter, 2009; Stoye, 2009; Kitagawa and Tetenov, 2018; Athey and Wager, 2021).

2 Institutional Background

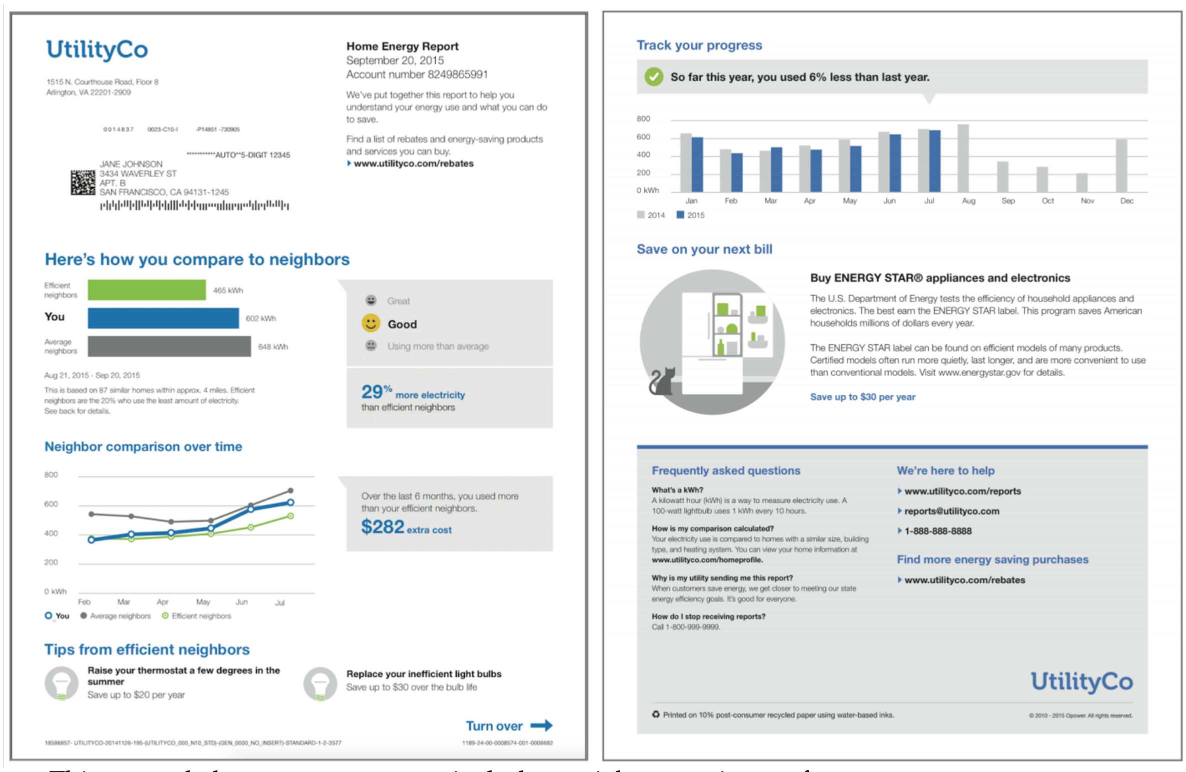
In this study, we use data on a home energy reports program administered by a major utility company in the Northeast United States. The program is administered by Opower, the leading provider of home energy reports in the United States. The program addresses imperfect information through information provision. It also leverages psychological effects of social comparisons to promote behavioral changes in energy usage (Schultz et al., 2007; Nolan et al., 2008).[4] These programs are motivated by previous findings that households may not be perfectly informed about energy costs, and that, even if they have the information, they may not devote attention to the issue. Thus, their energy consumption decisions might not be privately or socially optimal, contributing to the "energy-efficiency gap" (Jaffe and Stavins, 1994; Gerarden et al., 2017). The program is implemented as bimonthly, personalized letters mailed to residential households. The letters contain clear messaging on historical monthly usage, comparisons to usage by efficient neighbors, and personalized recommendations on energy conservation measures and energy-efficient products. Seminal studies on the effects of the program include Schultz et al. (2007), Nolan et al. (2008), Allcott (2011), Ayres et al. (2013), and Allcott and Rogers (2014).

For the utility company we study, the first cohort of Opower reports were issued in March 2013, and mailed to about 184,000 electricity account customers. Roughly 90 percent of the accounts enrolled in the program were randomly assigned into the treatment group, while the rest were assigned to the control group. From 2013 to 2018, there were seven additional "waves" of accounts added to the program. Electricity and gas waves were separately implemented, with different letters focusing on electricity and

[4]Comparisons often involve stating "Here's how you compare to neighbors," and "You use X percent more electricity than efficient neighbors."

gas consumption. After the initial enrollment, the treatment group received the Opower letters every other month; the control group did not receive letters. A sample Opower letter is shown in Figure 1. Unless customers explicitly opted out, they remained in the designated group after the initial assignment.

Figure 1: Sample Home Energy Report



Notes: This example home energy report includes social comparisons of energy usage, energy conservation tips, and a summary of historical usage. Some reports include additional content such as information about energy efficiency rebates. *Source:* Home Energy Reports Customer Service Guide. Oracle.

3 Data

This analysis utilizes three datasets: a four-year panel of household electricity consumption data; the treatment status of accounts enrolled in the Opower program; and cross-sectional data on household characteristics. All three datasets are provided by a major utility company in the Northeast United States. The data include all residential account holders of the utility in a single Northeast state.

3.1 Electricity Consumption Data

The first dataset contains information on monthly electricity consumption for residential electricity accounts from January 2014 to May 2018. We observe the billing account number, rate group, and the electricity consumption in kilowatt-hours (kWh) for each billing reading period. Electricity consumption and Opower program data are merged by account number. We trim observations with billing reading periods that deviate far from 30 days, keeping observations with reading periods between 25 and 34 days. After data cleaning, there are about 22 million observations in the sample, for about 420,000 unique billing accounts per month. The sample is unbalanced because some accounts existed for less than the entire sample period for a variety of reasons (e.g., the customer initiated or terminated service, incorrect readings, and missing readings). See Appendix A for additional details.

3.2 Opower Program Data

The second dataset contains Opower participation data by electricity account number. Table 1 shows the number of accounts for each wave. For each wave, 50 to 90 percent of accounts were randomly assigned treatment. The third, sixth, and seventh waves occurred during the sample time frame, and they are included in the sample for estimation. The first and second waves are excluded because they preceded the sample period for electricity consumption data, and therefore lack pre-treatment consumption data. The fourth wave only enrolled new account holders, so it also lacks pre-treatment consumption data. The fifth wave only included gas accounts. The eighth wave has insufficient post-treatment consumption data because it occurred too late in the period for which electricity consumption data are available.

Table 1: Number of accounts in each Opower wave from the start of the program to 2018.

Opower wave	Month/Year	Number of electric accounts assigned into Opower	Number of electric accounts treated
1	03/2013	183,789	166,911
2	04/2013	19,838	17,943
3	**03/2014**	**43,435**	**36,759**
4	08/2014	42,069	38,174
5	10/2015	0	0
6	**08/2016**	**25,974**	**12,992**
7	**03/2017**	**44,372**	**31,199**
8	02/2018	31,534	21,688

Notes: Wave 5 is a pure gas wave with no electric accounts. The **highlighted waves (3, 6, and 7)** have sufficient pre- and post-treatment data to be included in our analysis.

3.3 Household Demographic Data

The third dataset contains six household characteristics at the electricity account level: annual household income, number of household members, building size, house size, house age, and marital status.[5] Annual household income is available in $5,000 bins that range from $0-$5,000 to $145,000 and above. Building size and house size are in square feet. We censor house size at 5,000 square feet. House age represents the original year the house was built, and we censor it at 1850. For our analysis, we restrict attention to accounts for which all six characteristics are available, which is roughly half of the accounts in waves 3, 6, and 7. See Appendix A for details. Summary statistics for electricity consumption and household demographics in the final sample are shown in Table 2.

[5]For single-family homes, building size and house size are the same. For multi-family buildings, house size refers to the size of specific unit the household resides in.

Table 2: Summary statistics for the analysis sample

	Mean	Median	SD	Min	Max
Monthly electricity consumption (kWh)	505	394	384	0	2,705
Income ($)	66,104	55,000	43,985	5,000	150,000
Number of household members	2.29	2	1.62	1	8
Building size (ft^2)	4,655	2,999	11,146	262	232,146
Unit size (ft^2)	1,886	1,540	1,053	210	5,000
House Year Built	1947	1950	36	1850	2013
Married	.482	0	.5	0	1
Households	49,536				
Observations	2,186,105				

Notes: The analysis sample consists of households in waves 3, 6, 7 for which electricity consumption and all demographic variables are available. See Appendix A for details. The sample is an unbalanced panel with monthly electricity consumption constructed from meter readings. The demographic variables are cross sectional, common across the sample period for the same household.

4 Evidence of Heterogeneous Treatment Effects

In this section, we estimate average and heterogeneous treatment effects to motivate our formal analysis of the gains from targeting treatment.

4.1 Average Treatment Effects

We first estimate the average treatment effect of the program on electricity consumption, and compare our results to previous findings. We adopt a difference-in-differences design similar to that used in prior work (e.g., Allcott, 2011). The identifying assumption for difference-in-differences is parallel trends in the absence of treatment. Validity of this assumption is ensured by Opower's use of random assignment to determine treatment within each wave, as long as the randomization itself is successful. To assess randomization, we present balance tests by wave in Appendix B. Differences in mean pre-treatment electricity consumption between the treated and control groups within each wave are small in magnitude and statistically indistinguishable from zero. This is also true for all observable demographic variables. In addition, we plot average electricity consumption by treatment arm around the time of treatment in Figure G.1. Average electricity consumption appears virtually identical across groups in the year prior to treatment, providing evidence of parallel pre-trends.[6]

[6]Despite the apparent success of randomization within each wave, there is variation in pre-treatment consumption and other household characteristics *across* waves. This provides suggestive evidence of selection in the wave assignment process. See Appendix B for more detail.

We estimate average treatment effects using the equation

$$kWh_{iwt} = \beta_1 Opower_i + \beta_2 Opower_i \times Post_{wt} + X_i\gamma + \delta_{wt} + \varepsilon_{iwt}, \quad (1)$$

where the dependent variable is electricity consumption by household i in wave w and year-month t in kilowatt-hours. $Opower_i$ is an indicator for receiving the Opower letter, and $Post_{wt}$ is an indicator for months after treatment began.[7] X_i is a vector of household demographics including 12-month average pre-treatment consumption, income, building size, house size, house age, marital status, and number of household members. Wave-by-year-month fixed effects, δ_{wt}, absorb seasonal variation in electricity consumption that is common between the treated and control groups within a wave, but allowed to vary across waves.[8] The coefficient of interest is β_2. Estimates are shown in Table 3.

Table 3: Average treatment effects

	Dependent variable: Electricity Usage in kWh			
	(1)	(2)	(3)	(4)
	All waves	Wave 3	Wave 6	Wave 7
Opower × Post	-5.79**	-6.31	-0.96	-3.11
	(2.55)	(4.83)	(2.49)	(2.69)
Demographics	Yes	Yes	Yes	Yes
Baseline usage	Yes	Yes	Yes	Yes
Wave × year-month FE	Yes	Yes	Yes	Yes
Control mean	473	579	428	422
Households	49,536	22,915	11,114	15,507
Observations	2,186,105	1,167,703	504,889	513,513

* p<0.10, ** p<0.05, *** p<0.01

Notes: This table shows the estimation results for equation 1 on electricity consumption. Wave-by-year-month FE indicates fixed effects for the interaction of wave and calendar sample month. Control means are the mean monthly electricity consumption for the control group in each estimation sample. Standard errors are clustered at the household level and shown in parentheses.

Our main specification, in column 1, yields a pooled average treatment effect on the treated of -5.79 kWh/month, which is equivalent to a reduction of about 1.2 percent in

[7] Opower letters were consistently mailed to households in the treatment group every two months; therefore households received information treatment throughout the post-assignment period. However, we don't observe the specific timing of each mailing, and so we only estimate the treatment effect relative to when Opower treatment status was assigned.

[8] The interaction of wave and time fixed effects allows for the possibility that selection across waves might induce a correlation between electricity consumption and treatment status due to variation across waves in both average electricity consumption and treatment shares. We also estimate models separately by wave.

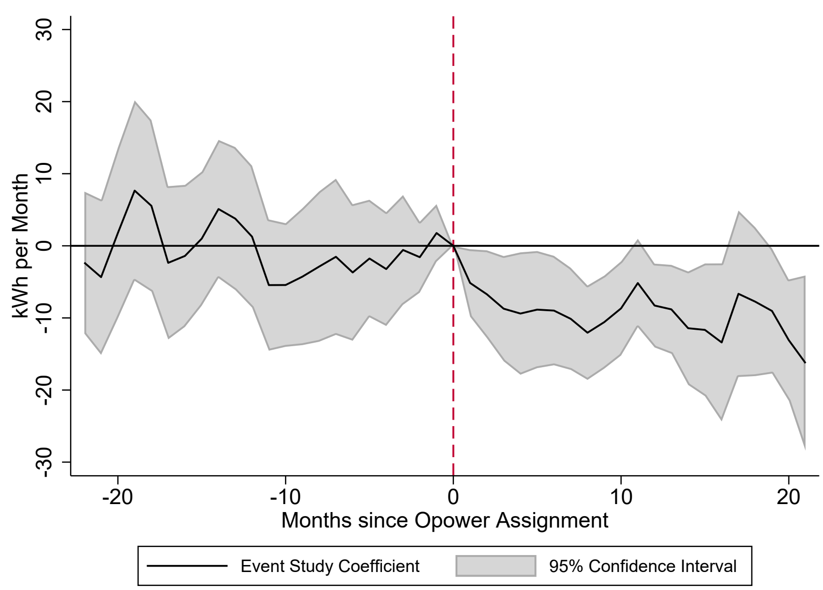
electricity consumption compared to the average pre-treatment level. This is in line with estimates from previous studies of home energy reports (e.g., Allcott, 2011; Allcott and Rogers, 2014). Columns 2 through 4 present estimates of wave-specific average treatment effects. The treatment effect point estimate is largest for wave 3 and smaller for later waves with lower baseline consumption levels. These estimates are robust to inclusion and exclusion of demographics and baseline electricity consumption (Appendix Table G.1).

We also use an event study framework to estimate and visualize dynamic treatment effects. Specifically, we regress electricity consumption on Opower treatment status interacted with event month relative to treatment and wave-by-year-month fixed effects:

$$kWh_{iwt} = \beta_1 Opower_i + \sum_{r=-23}^{r=-1} \mu_r \times Opower_{it}^r + \sum_{r=1}^{r=22} \mu_r \times Opower_{it}^r + X_i\gamma + \delta_{wt} + \varepsilon_{iwt}. \quad (2)$$

where $Opower_{it}^r$ is an Opower treatment indicator for being r months relative to i's assignment to treatment. All estimates are relative to the month of Opower assignment ($r = 0$). Figure 2 presents regression results in graphical form for the pooled sample. Wave-specific results are in Figure G.2.

Figure 2: Event study of the pooled sample



Notes: Treatment effect estimates by event month for the pooled sample. The month of Opower assignment (0) is normalized to zero. Observations with event month prior to -22 are grouped to event month -23, and those with event month after event 21 are grouped to event month 22; these endpoints are omitted from the plots. The endpoints are chosen based on the event months for wave 6, which was initiated close to the middle of sample. A time-invariant Opower treatment indicator, household characteristics, and wave-by-year-month fixed effects are included in the event study model but omitted from this plot.

There is no evidence of a violation of parallel pre-trends in Figure 2. After treatment the monthly electricity consumption of treated households is about 1 percent lower on average, relative to control households. After the treatment assignment in event month zero, treated households continued to receive letters with energy reports every two months. The treatment effects are persistent, with slight variations in magnitude, which is consistent with findings from Allcott and Rogers (2014). The lasting impact of the letters on households' behavior underscores the potential importance of using targeting to treat households that are likely to reduce consumption, and to avoid treating households that might increase consumption.

4.2 Heterogeneous Treatment Effects

We use parametric models in a triple-differences framework to provide suggestive evidence of heterogeneous treatment effects using the following estimating equation:

$$\begin{aligned} kWh_{iwt} = {} & \beta_1 Opower_i + \beta_2 C_i + \beta_3 Post_{wt} + \beta_4 Opower_i \times C_i + \beta_5 C_i \times Post_{wt} \\ & + \beta_6 Opower_i \times Post_{wt} + \beta_7 Opower_i \times C_i \times Post_{wt} \\ & + C_i \times X_i \gamma + C_i \times \delta_{wt} + \varepsilon_{iwt}, \end{aligned} \tag{3}$$

where C_i indicates the covariate of interest at the household level, discretized into groups above and below the median within wave. We interact covariate groups with wave-by-year-month fixed effects, $C_i \times \delta_{wt}$, to absorb seasonal variation in electricity use that is common between the treated and control groups within each covariate group within each wave. As before, we include a set of demographic variables X_i, but we interact these demographic variables with the covariate of interest C_i to allow for differences in the relationship between demographics and electricity consumption across covariate groups. The set of demographic variables in the regression excludes the covariate C_i.[9] The coefficients of interest are β_6 and β_7, which represent treatment effects for households below and above the median of each covariate we consider.

Table 4 summarizes the results. The estimates suggest that the effect of information provision is larger in magnitude for households with higher pre-treatment consumption, larger houses, higher incomes, and older homes.[10] The difference between groups is largest in magnitude for pre-treatment consumption. As a result, our targeting analysis

[9] C_i, $Post_{wt}$, and $C_i \times Post_{wt}$ are collinear with the $C_i \times \delta_{wt}$ fixed effects, and are therefore omitted from the regressions, but we include them in equation 3 for clarity.

[10] Our finding that treatment effects are larger for households with higher pre-treatment consumption is consistent with previous studies of electricity consumption (e.g., Allcott, 2011; Prest, 2020) and water consumption (e.g., Ferraro and Price, 2013; Ferraro and Miranda, 2013).

focuses on the effectiveness of treatment rules that combine pre-treatment consumption with each of the other covariates.

Table 4: Heterogeneous treatment effects

	Dependent Variable: Electricity Usage in kWh			
	Baseline Usage	House Size	Income	House Year Built
Opower × Post × Below Median	-2.65	-3.53	-2.20	-6.66*
	(2.26)	(3.41)	(3.55)	(3.68)
Opower × Post × Above Median	-12.45**	-8.64**	-9.08**	-4.89
	(6.17)	(3.76)	(3.58)	(3.53)
Demographics × category	Yes	Yes	Yes	Yes
Baseline usage × category	No	Yes	Yes	Yes
Wave × year-month × category FE	Yes	Yes	Yes	Yes
p-value, test of equal coefficients	0.14	0.31	0.17	0.73
Control mean	473	473	473	473
Households	49,536	49,536	49,536	49,536
Observations	2,186,105	2,186,105	2,186,105	2,186,105

* p<0.10, ** p<0.05, *** p<0.01

Notes: This table shows estimation results for the coefficients of interest from equation 3. Demographics × category and Baseline usage × category indicate the interaction of covariates, excluding the covariate of interest, with the above or below median indicator for the covariate of interest. Wave × year-month × category FE indicates fixed effects for the interaction of wave, sample month, and covariate level (above or below median). Estimates for all nuisance parameters in equation 3 are omitted for clarity. Control means are the mean monthly electricity consumption for the control group in each estimation sample. Standard errors are clustered at the household level and shown in parentheses. The p-values are from a Wald test of equal coefficients for the above- and below-median groups.

5 Empirical Strategy for Targeting

Given the evidence of heterogeneous treatment effects in the previous section, it may be possible to improve the cost-effectiveness of the program by identifying and treating households that respond most favorably to home energy reports. This section outlines our methodology for investigating the potential gains from this treatment targeting. Our approach integrates the process of searching for treatment rules that would maximize the net benefits of the program with estimation of those net benefits.

Our goal is to empirically select treatment assignment rules that maximize expected energy and cost savings, and to evaluate their performance.[11] A treatment rule, denoted by π, determines whether a household with a given set of characteristics is treated. The value of any candidate rule π is

[11] We primarily refer to these rules as "treatment rules" for brevity. These are also referred to as treatment assignment "policies" in the literature, and we occasionally use that terminology.

$$V(\pi) \equiv E\left[Y^1 \cdot 1\{X \in \pi\} + Y^0 \cdot 1\{X \notin \pi\}\right], \tag{4}$$

where Y^0 and Y^1 denote potential outcomes. The binary indicator variable $1\{X \in \pi\}$ is one for households with characteristics X that are treated by rule π, and zero otherwise. The expectation is taken with respect to the joint distribution of potential outcomes and characteristics for the population of interest (i.e., the "target population"). In words, the value of treatment rule π is simply the average of individual-level potential outcomes under that rule.

It is impossible to directly compute the value of each candidate rule due to the fundamental problem of causal inference: only one potential outcome is observable, not both. We utilize the empirical welfare maximization method introduced by Kitagawa and Tetenov (2018) to circumvent this challenge. This approach rests on three key assumptions. First, treatment must be independent of potential outcomes. Second, there must be overlap in the propensity score in the sampled population. Third, the joint distributions for the sampled population used for estimation and the target population must be equivalent. In our application, randomization ensures that the first two assumptions hold. Validity of the third assumption follows from the fact that our targeting analysis focuses on evaluating the gains from treatment in the same set of households that are used for estimation. Kitagawa and Tetenov show that under these three assumptions, equation 4 can be rewritten as

$$V(\pi) = E(Y^0) + E\left[\left(\frac{YD}{e(X)} - \frac{Y(1-D)}{1-e(X)}\right) \cdot 1\{X \in \pi\}\right], \tag{5}$$

where Y denotes observed outcomes, D denotes treatment status, and $e(X)$ denotes the propensity score $E[D|X]$. Thus, maximizing $V(\pi)$ is equivalent to maximizing the expected gain from treatment rule π relative to no treatment. Crucially, the gain from treatment rule π can be estimated using observed rather than potential outcomes. Furthermore, maximizing the gain from treatment rule π alone is isomorphic to maximizing $V(\pi)$, because $E(Y^0)$ is invariant to the choice of treatment rule.

Since home energy reports are intended to promote energy conservation, we recast this welfare-maximization problem as a minimization problem. Our goal is to learn optimal treatment rules ($\widehat{\pi}$) that minimize the sample analog of electricity consumption and cost relative to no treatment:

$$\widehat{\pi} \in \arg\min_{\pi \in \Pi} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left(\frac{Y_i D_i}{e(X_i)} - \frac{Y_i(1-D_i)}{1-e(X_i)} \right) \cdot \mathbb{1}_{(X_i \in \pi)} \right\}, \tag{6}$$

where Y_i denotes outcomes, D_i is the observed treatment status in the original experiment, and X_i are pre-treatment characteristics used for targeting. Π denotes the set of treatment rules under consideration. To account for non-random selection into waves, yet also take advantage of randomization within waves, we use the share of households treated within each wave as the propensity score $e(X_i)$ for all households in that wave.

We use multiple outcomes to evaluate the potential gains from targeted treatment. In the analyses that focus on energy conservation, Y_i is electricity consumption in kWh/month. In our other analyses, we focus on expected cost savings, so Y_i is a total cost measure that includes both energy costs and program implementation costs. To convert from observed electricity consumption to energy cost, we use two different prices: first, the retail electricity rate; and second, an estimate of the average short-run social marginal cost of electricity generation.[12] We include a monthly program cost of 76.5 cents per household based on administrative costs reported by the utility company.[13] In all the analyses, we use the difference between average electricity consumption in the year after treatment relative to the year before treatment to construct Y_i. We seek to minimize this difference, as it can take on positive or negative values, with negative values corresponding to reductions in energy consumption. The result is that our targeting approach exploits similar identifying variation to the difference-in-differences design in Section 4.

The characteristics we use for targeting are 12-month average pre-treatment electricity consumption, household income, house size, and house age. For computational simplicity, we limit our attention to two covariates at a time. In addition to the suggestive evidence from Section 4.2, prior studies have found pre-treatment consumption to be a significant predictor of treatment effects (e.g., Allcott, 2011). We thus focus on three pairs of variables: pre-treatment consumption and income; pre-treatment consumption and house size; and pre-treatment consumption and house age.[14]

To estimate optimal treatment rules, we restrict our attention to two types of rules. This has both theoretical and practical advantages. From a theoretical perspective, constraining the form of the treatment rules ensures that the expected loss from failing to adopt the

[12]The average rate for the standard-income residential sector in the sample period is \$0.177/kWh. Our estimate of the average short-run social marginal cost, \$0.065/kWh, is from Borenstein and Bushnell (2022).

[13]We do not observe the marginal cost of a home energy report, so we estimate it using the average cost. The average monthly program cost is calculated as the total annual program implementation expense (\$2,464,200) divided by the number of participants (268,263), divided by the number of months per year (12). Program implementation expenses and customer participation counts are reported in Table E-3 column 3 and Table E-1 column 8 of Public Utilities Commission Docket #4527. In principle, our analysis could be replicated using alternative implementation cost estimates to account for the possibility that this average cost measure is biased upwards relative to the true marginal cost due to fixed costs.

[14]Extending the analysis to additional covariates imposes a large computational burden in our application.

ideal rule converges to zero as the sample size increases. From a practical perspective, this reduces the computational burden of estimation. It also helps prevent overfitting.[15] Finally, restricting attention to a set of simple and transparent rules, as we do, could be appealing to utilities and their regulators.

The first type of rule we consider is a two-dimensional quadrant rule. This consists of all possible splits of a two-dimensional characteristic space into four quadrants, one of which is treated. This type of rule defines thresholds for the two covariates that determine treatment, so that its implementation is simple and transparent. Formally, the set of treatment rules we consider is

$$\Pi_Q \equiv \{s_1(X_1 - t_1) \geq 0 \quad \& \quad s_2(X_2 - t_2) \geq 0, \quad s_1, s_2 \in \{-1, 1\}, \quad t_1, t_2 \in \mathbb{R}\}. \tag{7}$$

Candidate treatment rules are described by four parameters: s_1, s_2, t_1, and t_2. The optimal rule is found via grid search of all possible combinations of grid points and orientations. Treatment is then assigned to the quadrant that minimizes expected electricity consumption and cost as defined in equation 6.

The second type of treatment rule we consider is a linear eligibility score. This allows for more flexibility in partitioning the characteristic space than quadrant rules. The specific set of treatment rules we consider is

$$\Pi^3_{LES} \equiv \left\{(\beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_1^3 + \beta_4 X_2 \geq 0), \quad \beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \in \mathbb{R}\right\}, \tag{8}$$

where one term (X_1) enters the eligibility score as a cubic function and the other term (X_2) enters linearly. The optimization problem in equation 6 can be tailored to this linear eligibility score as follows

$$\min_{\beta} \frac{1}{N} \sum_{i=1}^{N} \underbrace{\left(\frac{Y_i D_i}{e(X_i)} - \frac{Y_i(1 - D_i)}{1 - e(X_i)}\right)}_{g_i} \mathbb{1}\{X_i^T \beta \geq 0\}, \tag{9}$$

where $X_i^T \beta$ denotes the dot product that defines the linear eligibility score in equation 8. We introduce g_i as shorthand for each observation's contribution to the sample average.

We use IBM's CPLEX Optimizer to solve for the optimal linear rule with cubic terms. Specifically, we follow Kitagawa and Tetenov (2018) in converting the optimization problem defined in equation 9 to the following mixed-integer linear programming problem

[15] We do not use data-dependent methods to select the types of treatment rules we consider. However, the analysis in Section 6.5 uses separate samples to derive and evaluate treatment rules. Those targeted treatment rules outperform randomization in almost all cases, which mitigates concern about overfitting.

for computational efficiency:

$$\min_{z,\beta} \frac{1}{N}\sum_{i=1}^{N} g_i\, z_i \quad \text{s.t.} \quad \frac{X_i^T\beta}{c_i} < z_i \le 1 + \frac{X_i^T\beta}{c_i} \;\; \forall\, i \quad , \quad z_i \in \{0,1\} \tag{10}$$

where z_i is a binary indicator for treatment assignment and β is a vector of "weights" on each covariate used to determine z_i. The constraints for z_i require that $z_i = 1$ if and only if $X_i^T\beta \ge 0$. c_i are constants greater than the suprema of $|X_i^T\beta|$. Appendix C provides additional detail on the process of solving for optimal treatment rules.

The estimation procedures for both types of treatment rule yield estimates of the optimal treatment assignment for each household, as well as estimates of the reductions in energy use and cost those treatment assignments would engender. These energy use and cost reductions can be expressed per household in the sample or on an aggregate basis. In both cases, the estimates of energy and cost reductions are relative to a baseline of no treatment.

We evaluate these estimated energy and cost savings relative to the savings from the actual treatment assignment in the original randomized controlled trial (RCT). We compute the savings from the actual RCT assignment using the inverse-probability weighting estimator of the average treatment effect on the treated: $\frac{1}{\sum D_i}\sum\left(Y_iD_i - \frac{Y_i(1-D_i)e(X_i)}{1-e(X_i)}\right)$. Since Y_i is defined as the difference between post-treatment and pre-treatment outcomes and we use experimental treatment shares for $e(X_i)$, this is equivalent to non-parametric difference-in-differences (see, e.g., Abadie, 2005). We multiply the average treatment effect on the treated by the share of households treated in the experiment, so that the results reflect savings per household in the sample rather than per household treated. Similarly, the estimated impact of each EWM rule is in terms of average energy or cost change per household in the sample, not per treated household. The number of households in the sample is the same for all EWM rules, because we use the same estimation sample for all analyses of the pooled sample.[16] This enables internally consistent comparisons between the RCT and the EWM rules.

Finally, to conduct inference, we use two complementary approaches. First, we bootstrap asymptotically valid confidence intervals for the effects of the estimated EWM rules in the population following the procedure outlined in Kitagawa and Tetenov (2018). To ensure adequate coverage, this approach searches over all candidate rules to identify the rules that lead to the biggest differences in savings between each bootstrap sample

[16] For the analysis of targeting using historical data in Section 6.5, we separate the pooled sample by wave, and estimate EWM rules by wave, with the results in terms of average savings per household in that wave.

and the original sample. This yields very conservative confidence intervals for the savings from EWM rules.

Second, to demonstrate the practical value of our approach, we compute the savings from applying the *specific* EWM rule we estimate from the original data, and evaluate those savings relative to the RCT. We construct confidence intervals for these estimates via bootstrap. This approach does not account for the effect of sampling variation on the specific EWM rule we estimate from the original data. This is similar in spirit to a great deal of empirical economic research, which selects a model for estimation and conducts inference based on that model without formally considering model specification as a source of uncertainty. However, these confidence intervals are conservative insofar as we apply the specific EWM rule estimated from the original data to each bootstrap sample, rather than searching over all candidate rules to optimize targeting within each bootstrap sample. We emphasize this second approach to inference throughout the paper. Appendix C provides more detail on both approaches.

6 Results

This section summarizes the performance of targeting the assignment of home energy reports to achieve three different objectives: minimizing energy consumption, minimizing private cost, and minimizing social cost.[17] Each analysis follows the same approach but yields different results due to the different objectives under consideration. Taken together, the results demonstrate the overall performance of targeting and highlight how that performance will depend in general upon the choice of objective. In addition, we demonstrate the utility of our approach by using pre-treatment electricity consumption on its own, without any additional household characteristics, to design targeted treatment assignment rules. Finally, we study how *ex-ante* targeting based on historical data compares to our main results, which devise targeting strategies *ex-post* based on all the data we observe. These results underscore the practical value of our approach.

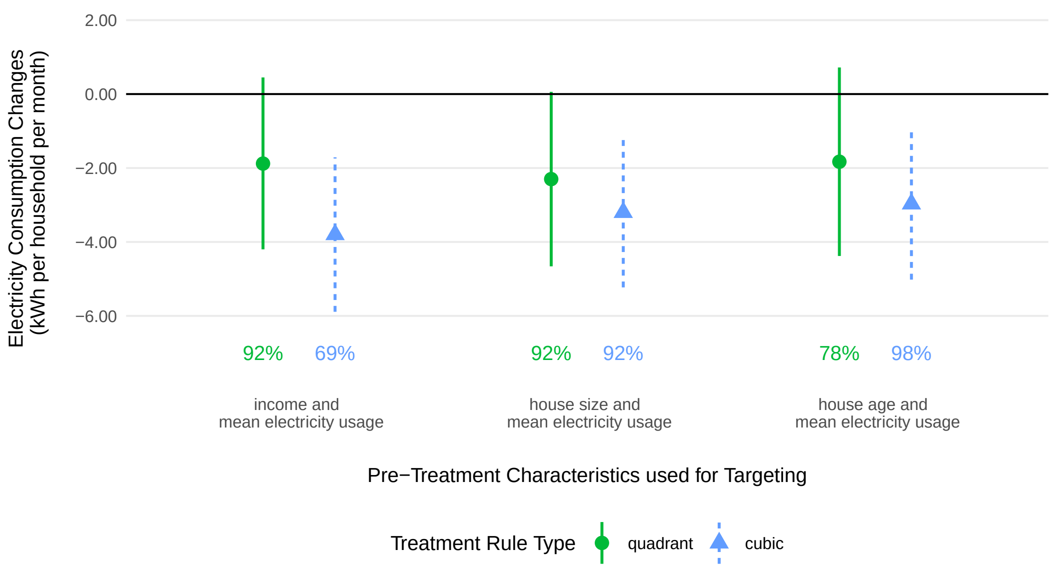
6.1 Energy Savings from Targeting

Figure 3 summarizes the energy savings that would be achieved by targeted treatment rules, relative to the energy savings from the original experiment as implemented. These

[17] Private cost savings are calculated by valuing electricity consumption at the retail electricity rate. Social cost savings are based on valuing electricity consumption at the short-run social marginal cost of electricity generation. See Section 5 for more details.

energy savings – and all other results – are normalized by the total number of households in the sample to facilitate direct comparisons between rules that treat different numbers of households. Table G.2 presents more detailed results, including estimates of the gains from targeting relative to a baseline of no treatment, and the share of households treated by each rule.

Figure 3: Estimates of the energy savings from targeted treatment assignment



Notes: Points indicate estimates of the energy savings that would be achieved from targeted treatment assignment, relative to the original experiment as implemented. Lines represent 95% confidence intervals. Quadrant rules are shown in green solid lines and cubic rules are shown in blue dashed lines. The percentage beneath each estimate represents the share of households that would be treated by that rule.

Figure 3 shows that targeted treatment assignment could reduce energy consumption beyond what was achieved through randomization. Point estimates suggest that targeting could double the energy savings that were generated by the program as implemented, which reduced energy consumption by 3.6 kWh per household per month. In total, targeting using this approach could result in electricity consumption reductions of more than three gigawatt-hours per year for the sample.[18]

Because this initial analysis focuses on energy savings and omits implementation costs, the rules treat all households that decrease consumption after treatment, and avoid treating households that increase consumption after treatment (on average across households, conditional on the characteristics used for targeting). On net, these targeted treatment rules tend to treat more households than the original experiment, which can be

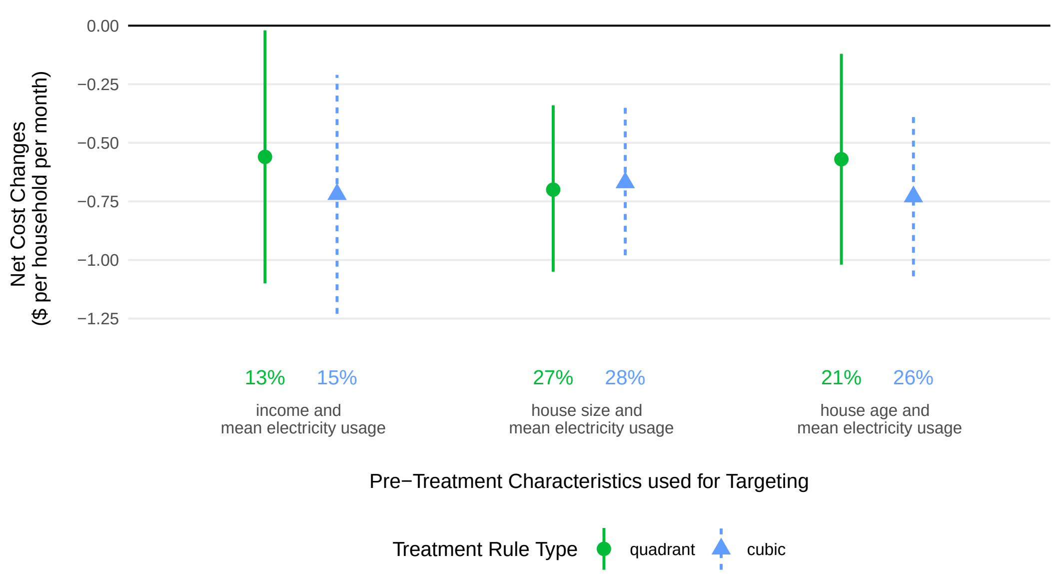
[18]This aggregate reduction in annual energy usage is relative a baseline of no treatment. It is calculated by multiplying point estimates from Table G.2 by the number of unique accounts in the sample as reported in Table A.1 and the number of months per year (12).

shown by comparing the shares that would be treated in Figure 3 to the benchmark of 72 percent treated in the original experiment. Compared with universal treatment, on the other hand, targeting excludes households that increase consumption in response to the letters, treating fewer households but achieving greater total energy savings.[19]

6.2 Private Cost Savings from Targeting

Figure 4 summarizes the potential gains from targeting treatment assignment to minimize electricity expenditures. This analysis values electricity conservation at the retail price of electricity and accounts for program implementation costs, as described in Section 5.[20] Additional results are presented in Table G.3.

Figure 4: Estimates of the private cost savings from targeted treatment assignment



Notes: Points indicate estimates of the private cost savings that would be achieved from targeted treatment assignment, relative to the original experiment as implemented. Lines represent 95% confidence intervals. Quadrant rules are shown in green solid lines and cubic rules are shown in blue dashed lines. The percentage beneath each estimate represents the share of households that would be treated by that rule.

We find that targeted treatment assignment could reduce electricity expenditures

[19]To be internally consistent, we compare potential energy savings from targeting to the average treatment effect computed using equation 6, which yields a reduction of 4.6 kWh per household per month from universal treatment.

[20]We refer to these expenditures, which are net of program implementation costs, as "private costs" throughout the text in order to distinguish them from analyses that focus on alternative outcomes. These are not comprehensive measures of consumer welfare because they omit costs incurred by households to reduce electricity consumption, as well as any direct impacts that receiving home energy reports may have on consumer welfare (both of which are unobserved). In principle, these costs could be estimated and incorporated into a targeting analysis using methods or results from Allcott and Kessler (2019).

beyond what the original experiment achieved. The six treatment rules in Figure 4 would reduce the cost of electricity consumption by 56 to 73 cents per household per month relative to the original experiment, based on point estimates. Furthermore, these estimates are statistically significant. Targeting outperforms universal treatment in terms of cost reduction, too. This is because targeting allows program administrators to avoid treating households for which the cost of distributing the letters is greater than the savings that would be achieved. The share of households that would be treated varies across rules, but overall fewer than 30 percent of the households would be treated, whereas 72 percent of households were treated by the original experiment. Moreover, compared with the rules that maximize energy conservation in Section 6.1, the rules that maximize private cost savings would treat fewer households.

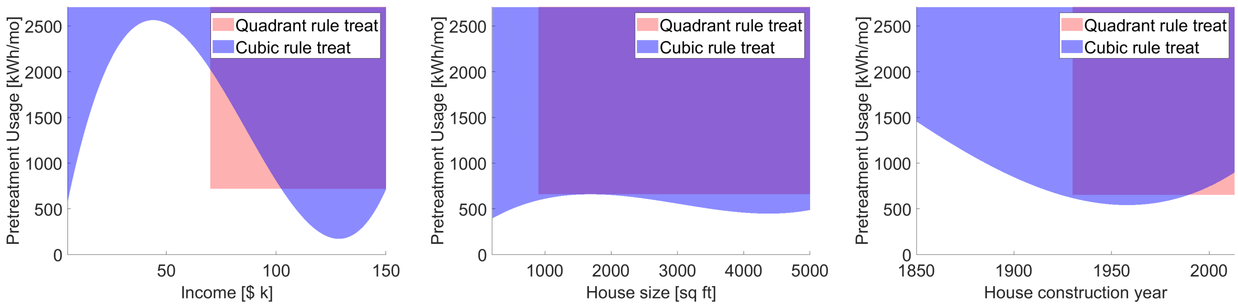
In absolute terms, these rules would achieve 65 to 82 cents in private cost savings per household per month, net of implementation costs. This is equivalent to roughly $385,000 to $485,000 of total net cost reductions per year for the sample. By contrast, we estimate that the original experiment generated net cost reductions of only $50,000. These results imply that targeting could increase the net cost savings from the program we study by an order of magnitude.[21]

Each rule we consider yields distinct treatment assignments. Figure 5 summarizes them graphically. The quadrant rules assign households to treatment based on thresholds for each covariate. For example, when using pre-treatment consumption and income to maximize private cost savings, households with average pre-treatment consumption above 720 kWh/month and incomes above $70,000 are treated. Figure 5a presents a visualization of this rule, with the treated quadrant in red. Cubic rules offer more flexibility in selecting treatment assignment boundaries on both covariate dimensions. The optimal cubic rule based on pre-treatment consumption and income is also plotted in Figure 5a. Like the quadrant rule, the cubic rule treats households with high incomes and high baseline consumption. However, it also treats some households with lower income but higher pre-treatment consumption, and vice versa.

Figures 5b and 5c present analogous plots for the other two sets of characteristics we consider. In Figure 5b, the optimal EWM quadrant rule treats households with larger houses and higher pre-treatment consumption. The cubic rule is very similar, but it treats households with homes of all sizes, with a cutoff for pre-treatment consumption that varies with house size. In Figure 5c, the optimal quadrant rule targets households with newer homes and higher pre-treatment consumption. The cubic rule also treats older

[21] Annual savings estimates are computed by multiplying point estimates from Table G.3 by the number of unique accounts in the sample as reported in Table A.1 and the number of months per year (12).

Figure 5: EWM rules for maximizing private cost savings



(a) Household income

(b) House size

(c) House age

Notes: EWM treatment assignment as a function of pre-treatment electricity consumption (y-axis) and three other household characteristics (x-axis), one for each plot. Households with characteristics that fall in shaded regions are treated by the quadrant rule (red) and cubic rule (blue).

homes, although only those with sufficiently high pre-treatment consumption levels.

6.3 Social Cost Savings from Targeting

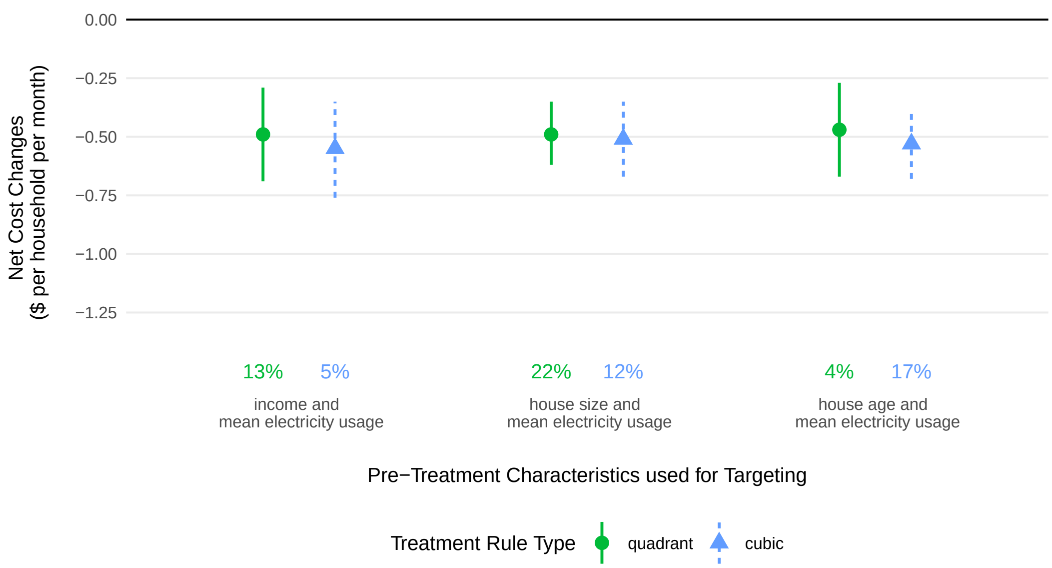
Our main analysis in Section 6.2 assesses the benefits of targeting when electricity conservation is valued at the retail rate. However, retail electricity rates often differ from the social marginal cost of electricity generation due to unpriced environmental externalities and the use of marginal rates to recover fixed costs, among other reasons. Therefore, we also investigate how targeted treatment assignment performs when valuing electricity savings at an estimate of the average short-run social marginal cost of electricity generation.[22]

We find that targeted treatment assignment would outperform the original experiment when valuing electricity consumption at its social cost. The results are summarized in Figure 6. Across the six rules we consider, targeting would achieve social cost savings of 47 to 55 cents per household per month. These savings are similar in magnitude to the results from valuing electricity at the retail rate in Section 6.2.

However, the *absolute* social cost reductions from targeted treatment assignment are lower when electricity consumption is valued at its social marginal cost than at the retail

[22]We use the term "social cost savings" to distinguish this analysis from our other analyses. These cost savings are not a comprehensive measure of the social welfare impacts of home energy reports. Because the utility only observes electricity consumption on a monthly basis, we are unable to use high-frequency cost measures to account for the covariance between social marginal cost and electricity consumption within a month. It is possible that alternative rules could outperform the rules we estimate after accounting for this intra-month variation. This could occur, for example, if households with certain demographics are more likely to conserve energy during times of peak electricity demand when the social cost of electricity generation is high. In addition, as mentioned in Section 6.2, our analysis does not account for unobserved costs incurred by homeowners who receive home energy reports.

Figure 6: Estimates of the social cost savings from targeted treatment assignment



Notes: Points indicate estimates of the social cost savings that would be achieved from targeted treatment assignment, relative to the original experiment as implemented. Lines represent 95% confidence intervals. Quadrant rules are shown in green solid lines and cubic rules are shown in blue dashed lines. The percentage beneath each estimate represents the share of households that would be treated by that rule.

electricity rate. This is because the social marginal cost of electricity is roughly one-third of the retail rate, so energy conservation is less valuable, while the implementation cost remains unchanged. As a result, the optimal treatment rules treat fewer households than rules that target private cost savings. This can be seen by comparing the share to treat labeled for each rule in Figures 4 and 6. The treatment rules themselves are similar to the rules that maximize cost savings when electricity usage is valued at the retail electricity rate, but with a higher threshold for pre-treatment consumption.

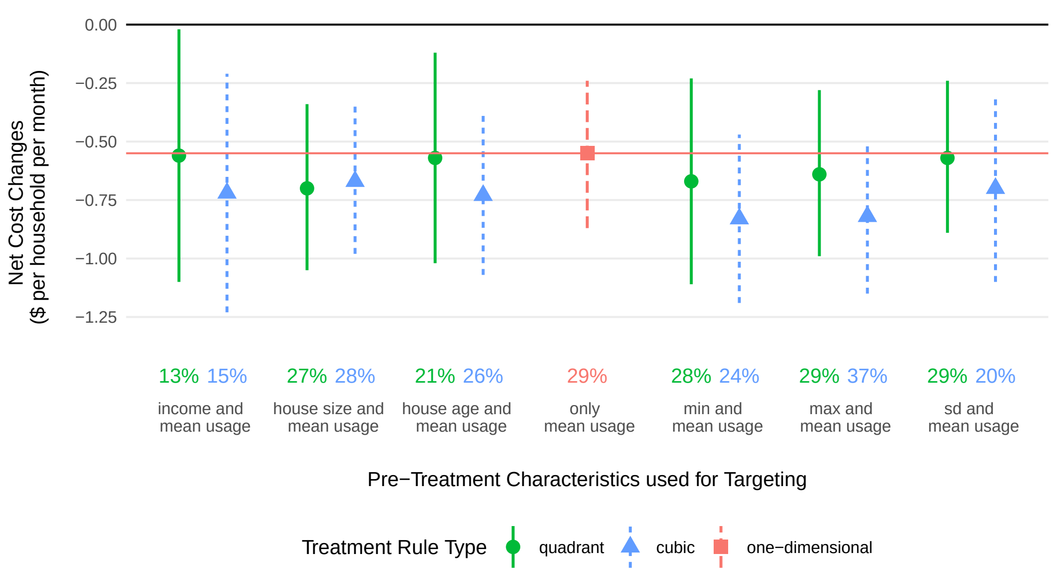
6.4 Targeting using Pre-treatment Consumption Data Only

One potential shortcoming of the preceding analysis is that it relies on detailed demographic data that may not be available for all households. To explore the potential of targeting based on pre-treatment consumption data alone, we derive a one-dimensional EWM treatment rule based on average pre-treatment consumption. We also consider two-dimensional EWM treatment rules based on average pre-treatment consumption and each of the following: minimum, maximum, and standard deviation of monthly pre-treatment consumption for each household.

We find that treatment rules based on pre-treatment consumption alone would achieve greater energy and cost savings than the treatment assignment used by the actual program.

Figure 7 plots the estimated private cost savings from targeting using these EWM rules relative to the program as implemented. It also includes the estimates based on household demographics from Figure 4. The cost savings that would result from targeting based on pre-treatment consumption alone are remarkably similar to the savings that would result from rules that incorporate other household characteristics. Even the one-dimensional EWM rule would achieve a level of savings that is similar to the savings that would be achieved by the quadrant rules that also utilize income and house age for targeting. Figure G.3 presents analogous results for energy savings and social cost savings, which are qualitatively similar to those in Figure 7.

Figure 7: Comparison of the gains from targeting with and without using demographic characteristics



Notes: Points indicate estimates of the **private cost savings** that would be achieved from targeted treatment assignment, relative to the original experiment as implemented. Lines represent 95% confidence intervals. Quadrant rules are shown in green solid lines, cubic rules are shown in blue dashed lines, and the one-dimensional rule is shown in red with long dashes. The percentage beneath each estimate represents the share of households that would be treated by that rule. The left side of the plot shows EWM rules based on both pre-treatment electricity consumption and demographic characteristics. The right side shows EWM rules based on pre-treatment electricity consumption only. To facilitate comparisons, a horizontal line indicates the savings associated with the one-dimensional EWM rule that uses only average pre-treatment electricity consumption for targeting.

Figure G.4 visualizes the estimated rules for private cost minimization. Figure G.4a highlights that targeting based on average pre-treatment consumption alone yields a cutoff rule that is similar to earlier results. That is, the optimal one-dimensional rule treats all households with baseline consumption above 640 kWh/month. In addition, Figures G.4b, G.4c, and G.4d illustrate how order functions of pre-treatment electricity consumption might capture additional information that is useful in guiding treatment rules.

These results demonstrate the potential for deriving targeting rules that only require monthly electricity consumption data. This is advantageous because data on household characteristics are often only available for a subset of utility customers – in our sample, less than half of accounts – whereas electricity consumption data are available for the universe of accounts. Using pre-treatment consumption data alone avoids the need to limit attention to a subset of utility accounts, as in our analysis, or to impute missing demographic data, as in Knittel and Stolper (2019). Expanding the analysis to all accounts would allow utilities to consider sending home energy reports to any of their customers. Indeed, when we extend this analysis to include all utility customers, we find that total private cost savings could be greater than \$790,000, or 63 percent higher than the savings that would be achieved by restricting attention to households for which demographic data are available.[23] In addition, avoiding imputation circumvents the possibility of imputation bias; it also circumvents the need to quantify the effect of uncertainty in imputed data on the estimated gains from targeting when conducting inference. Given that we find it is possible to achieve similar benefits to more complicated targeting rules using pre-treatment electricity consumption data alone, our finding underscores the practical value of targeting in this setting.

6.5 Targeting using Historical Data

The previous analyses serve as a proof of concept. We used experimental data on all the households that we can observe to design treatment rules for those same households. In practice, utilities only have access to historical data to design energy conservation programs. To mimic this reality, this section studies the performance of the EWM method when using past experimental waves to derive treatment rules for future waves. This provides a policy-relevant benchmark for implementation of the method.

One complication is that the three home energy report waves differ on observable characteristics. This is evident when comparing means for each characteristic across Tables B.1, B.2, and B.3. The EWM method can be extended to estimate treatment rules in situations where the characteristics of sample and target populations differ, but have common support. This can be done by reweighting the original EWM problem as follows:

$$\widehat{\pi}^T \in \arg\min_{\pi \in \Pi} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left(\frac{Y_i D_i}{e(X_i)} - \frac{Y_i(1-D_i)}{1-e(X_i)} \right) \cdot \rho(X_i) \cdot \mathbb{1}_{(X_i \in \pi)} \right\}, \tag{11}$$

[23] This sample contains 106,582 households, which is more twice the size of pooled sample that we use throughout the paper.

where $\rho(x) \equiv P_X^T(x)/P_X^S(x)$ is the density ratio of the marginal distributions of X for the target and sample populations. In this case, the sample population is a past wave of N households for which the covariates X_i, treatment status D_i, and the outcome Y_i are observable. The target population is a future wave for which only the covariates X_i are available because treatment has not yet occurred. We implement this approach as follows:

1. We estimate the density ratios $\rho(x)$ as the ratio of the empirical probability densities of the sample wave and the target wave. Specifically, we compute each density $P_X^w(x)$ as the number of households at each unique covariate level divided by the total number of households within the wave, and then take ratios: $\rho^{36}(x) = \frac{P_X^6(x)}{P_X^3(x)}$ and $\rho^{67}(x) = \frac{P_X^7(x)}{P_X^6(x)}$.

2. For each target wave, we reweight g_i as described in Section 5 using the density ratios to form $g_i' = \left(\frac{Y_i D_i}{e(X_i)} - \frac{Y_i(1-D_i)}{1-e(X_i)}\right) \cdot \rho(X_i)$, which represents the contribution of sample wave household i to the expected effects of the program in the target wave.

3. We use the reweighted g' to estimate EWM treatment rules by searching over candidate rules as before.

4. Finally, to mimic the real-world implementation of this approach, we use data from the actual experiment on the target wave to compute the gains from targeting. This yields estimates of the *ex-post* savings that would have been achieved had the treatment rule derived using data from the earlier wave been applied to the later wave in practice. We then compare these estimated savings with the savings from the original experiment. The savings estimates are all normalized to reflect savings per household in the target wave, enabling direct comparisons across rules that treat different numbers of households.

Table 5 summarizes results for the cubic rules, with all figures estimated relative to treatment assignment in the original experiment. The findings show that the cubic rules based on historical data would generate more energy savings than were generated by the program as implemented. In fact, in some cases, the energy savings are likely to be more than double the energy savings from the program that was put in place. The rules designed to minimize private and social costs also tend to outperform the original treatment assignment. In the cases where the cubic rules do not reduce costs relative to the original experiment, the difference in cost is small in magnitude.

This analysis of targeting using historical data provides some insight into the out-of-sample performance of the EWM method, since the treatment rules were derived

Table 5: Estimates of the gains from targeting using historical data

Target wave	Sample wave	Pre-treatment characteristics used for targeting	Energy changes kWh/hh-month	Private cost changes $/hh-month	Social cost changes $/hh-month
6	3	Income and mean usage	-0.50	-0.06	-0.22
		House size and mean usage	0.08	0.13	-0.01
		House age and mean usage	-1.80	-0.33	-0.23
7	6	Income and 21mean usage	-0.17	0.00	-0.41
		House size and mean usage	-2.42	-0.11	-0.39
		House age and mean usage	1.37	-0.08	-0.41

Notes: Net energy and cost changes are calculated by applying the EWM rule estimated on the sample wave to data from the target wave. These results are point estimates of the *ex-post* savings that would have been achieved had the treatment rule estimated on the earlier wave been applied to the later wave in practice. Negative values indicate reductions in energy consumption and cost, all of which are relative to the outcomes attained under the program's actual treatment assignment. All EWM rules are linear rules with cubic terms.

and evaluated on two different samples. For comparison, we present evidence on the in-sample performance of the EWM method by learning and evaluating treatment rules on the same sample wave in Appendix D. The results show that targeting *ex-post* is more effective than targeting using historical data, which could be due to differences in households across waves, overfitting, or a combination of the two.[24]

While the EWM rules estimated using reweighted historical data do not achieve the performance of the EWM rules estimated using target wave data by construction, this analysis provides a proof of concept that it is possible to outperform randomization in practice. This approach could be implemented in other settings in which treatment is already staggered over time (for example, due to budget constraints). Or, treatment could be staggered over time to use information from one or more initial experiments to improve targeting of a larger program implemented later in the spirit of adaptive experimental design (e.g., Kasy and Sautmann, 2021).

7 Discussion

In this study, we find that using empirical welfare maximization to target treatment assignment could significantly improve the effectiveness of home energy report programs.

[24]We explore the possibility of overfitting further in Appendix D. Specifically, we perform a hold-out analysis on the pooled sample to assess whether the EWM method is prone to overfitting in this context. This approach ensures that the training and testing data are drawn from the same distribution, enabling us to isolate overfitting from differences in households across waves. We find that one-dimensional rules based on average pre-treatment electricity consumption tend to outperform more flexible rules on average. The results suggest that more flexible rules are susceptible to overfitting. Overall, however, the EWM method outperforms randomized treatment assignment and no treatment, both in and out of sample.

Two other recent studies have conducted similar analyses using an alternative approach referred to as "plug-in rules." This method consists of predicting conditional average treatment effects for all households, and assigning treatment to households with predicted savings. Allcott and Kessler (2019) implement this approach; they find that using a rule that treats all households with above-median predicted savings increases natural gas conservation by 85 percent relative to the original experiment. Knittel and Stolper (2019) take a similar approach, using causal forests to predict household-specific treatment effects. They find that targeted treatment assignment would achieve a private cost reduction of \$1.17 per household per month and a social cost reduction of \$0.26 per household per month relative to no treatment.[25] The magnitudes of these energy and cost savings are similar to our results.

Compared with plug-in rules, our approach has four main advantages. First, the EWM method we employ has desirable theoretical properties: it achieves minimax optimal rates of convergence for utilitarian regret in finite samples (Kitagawa and Tetenov, 2018). By contrast, the use of plug-in rules is often motivated by asymptotic approximations (Manski, 2021), and we are not aware of results that prove finite-sample optimality of plug-in rules generically.

Second, the EWM method produces transparent rules for assigning households to treatment. Plug-in rules that use complex models for predicting treatment effects may be less transparent. In our setting, the gains from more complex treatment assignment rules based on observable characteristics may be modest relative to simple treatment rules. This is because targeting based on average pre-treatment electricity consumption alone performs well, as discussed in Section 6.4.

Third, the EWM method can accommodate constraints on the types of policies under consideration. For example, the approach can require or restrict targeting along certain dimensions. While this is not a focus of our application, these features are advantageous in regulatory settings that may require transparency or impose constraints on targeting based on demographics. Another example is that the method can be applied in contexts with exogenous budget constraints. We examine the performance of the EWM method under such budget constraints in Appendix E.

Finally, the EWM method integrates the decision problem and statistical inference. Energy and cost savings are explicitly maximized in the process of deriving optimal treatment assignment rules, and valid inference procedures for these savings exist. By

[25] Using numbers reported in Knittel and Stolper (2019), average private cost reduction per household-month is calculated as the reported total gain of \$6.3M/year divided by the reported sample of 449,824 households and 12 months. Average monthly social cost reduction per household-month is similarly calculated using the reported welfare gain of \$1.4M/year.

contrast, the plug-in approach consists of estimating conditional average treatment effects and using them to determine treatment assignment in two separate procedures. Furthermore, the availability of valid inference procedures depends on the estimation methods used in any given application of the plug-in rule approach.

8 Conclusion

This paper investigates the potential for targeted treatment rules to improve the net benefits from a large-scale behavioral intervention to encourage household energy conservation. We are able to derive simple and transparent treatment rules that would produce significant improvements in program performance. Our estimates suggest that the reduction in electricity consumption from targeted treatment assignment could double the consumption reduction achieved by the randomized treatment assignment used in the program we study. Furthermore, we estimate that targeted treatment assignment could achieve cost savings that are an order of magnitude larger than the savings achieved by the original program. Our results have clear policy implications: it is possible to significantly improve the performance of future home energy report programs through targeted treatment.

One limitation of our analysis is that we cannot directly optimize social welfare due to data constraints. We only observe electricity consumption on a monthly basis, while the social costs of electricity consumption vary considerably within a month. In addition, we do not observe all costs incurred by households who receive home energy reports. We are not aware of any targeting analysis that accounts for both of these potential sources of inefficiency.

A second, related limitation is that our analysis focuses on aggregate energy consumption and its economic implications. Targeting may have important distributional impacts. For example, if targeting restricts treatment to high-income households, it may provide benefits to them without providing similar benefits to low-income households.[26] In Appendix F we document how targeted treatment assignment would affect households of different incomes and races. We find that simple rules based on pre-treatment electricity consumption and household income treat high-income households at a disproportionately high rate. However, rules that condition on other characteristics lead to more equitable

[26] Whether home energy reports provide net benefits to households is an empirical question. Allcott and Kessler (2019) find heterogeneity in households' willingness to pay for home energy reports, and find that it is positively correlated with household income. We observe neither households' willingness to pay for home energy reports nor the costs households incur to reduce energy consumption after receiving them, so we are unable to conduct a comprehensive analysis of the distributional impacts of targeting.

treatment assignments. While we do not observe the race of each household, we compare different treatment assignment rules at the zip code level and do not find strong evidence of racial differences in treatment assignment. Further analysis of the distributional impacts of targeting is an important area for future research.

As a final caveat, the treatment rules and quantitative results of our analysis are specific to the context we study. We do not claim that the targeted treatment rules we derive for electricity conservation in a Northeast state would necessarily generate similar outcomes in other states or for other markets such as natural gas. For one, electricity rates in the state we study are among the highest in the country. These high electricity rates may have led to capital investment or behavior that affect how households respond to home energy reports, both on average and across the distribution. Households may also respond differently to home energy reports due to differences in consumer preferences across states. In addition, the benefits associated with energy conservation vary across states due to variation in the short-run social marginal cost of electricity generation. For these and other reasons, the gains from targeting in other states may be smaller or larger than the gains we estimate in this paper. While the specific treatment rules and estimated benefits are unique to our setting, the approach we take is generic. Our contribution is not to develop a specific treatment rule to apply broadly, but is instead to demonstrate the potential gains from applying a novel method to derive targeted treatment rules.

Using statistical treatment rules to optimize program design is receiving growing attention in economics, medicine, and many other disciplines. We contribute to this literature by applying empirical welfare maximization to optimize the design of a large-scale behavioral energy conservation program. Our results underscore the practical value of these methods and their potential to generate significant benefits in many domains.

References

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, 72(1):1–19.

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(9-10):1082–1095.

Allcott, H. (2015). Site selection bias in program evaluation. *The Quarterly Journal of Economics*, 130(3):1117–1165.

Allcott, H. and Kessler, J. B. (2019). The welfare effects of nudges: A case study of energy use social comparisons. *American Economic Journal: Applied Economics*, 11(1):236–76.

Allcott, H., Knittel, C., and Taubinsky, D. (2015). Tagging and targeting of energy efficiency subsidies. *American Economic Review*, 105(5):187–91.

Allcott, H. and Mullainathan, S. (2010). Behavior and energy policy. *Science*, 327(5970):1204–1205.

Allcott, H. and Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review*, 104(10):3003–37.

Athey, S. and Wager, S. (2021). Policy learning with observational data. *Econometrica*, 89(1):133–161.

Ayres, I., Raseman, S., and Shih, A. (2013). Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. *The Journal of Law, Economics, and Organization*, 29(5):992–1022.

Borenstein, S. and Bushnell, J. B. (2022). Do two electricity pricing wrongs make a right? cost recovery, externalities, and efficiency. *Forthcoming in American Economic Journal: Economic Policy*.

Brandon, A., Ferraro, P., List, J. A., Metcalfe, R., Price, M., and Rundhammer, F. (2017). Do the effects of social nudges persist? Theory and evidence from 38 natural field experiments. Working paper.

Brandon, A., List, J. A., Metcalfe, R. D., Price, M. K., and Rundhammer, F. (2019). Testing for crowd out in social nudges: Evidence from a natural field experiment in the market for electricity. *Proceedings of the National Academy of Sciences*, 116(12):5293–5298.

Byrne, D. P., Nauze, A. L., and Martin, L. A. (2018). Tell me something i don't already know: Informedness and the impact of information programs. *Review of Economics and Statistics*, 100(3):510–527.

Christensen, P., Francisco, P., Myers, E., Shao, H., and Souza, M. (2021). Energy efficiency can deliver for climate policy: Evidence from machine learning-based targeting. Working paper.

Costa, D. L. and Kahn, M. E. (2013). Energy conservation "nudges" and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association*, 11(3):680–702.

Davis, L. W., Fuchs, A., and Gertler, P. (2014). Cash for coolers: Evaluating a large-scale appliance replacement program in mexico. *American Economic Journal: Economic Policy*, 6(4):207–38.

Ferraro, P. J. and Miranda, J. J. (2013). Heterogeneous treatment effects and mechanisms in information-based environmental policies: Evidence from a large-scale field experiment. *Resource and Energy Economics*, 35(3):356–379.

Ferraro, P. J. and Price, M. K. (2013). Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment. *Review of Economics and Statistics*, 95(1):64–73.

Fowlie, M., Greenstone, M., and Wolfram, C. (2018). Do energy efficiency investments deliver? Evidence from the weatherization assistance program. *The Quarterly Journal of Economics*, 133(3):1597–1644.

Gerarden, T. D., Newell, R. G., and Stavins, R. N. (2017). Assessing the energy-efficiency gap. *Journal of Economic Literature*, 55(4):1486–1525.

Hirano, K. and Porter, J. R. (2009). Asymptotics for statistical treatment rules. *Econometrica*, 77(5):1683–1701.

Jaffe, A. B. and Stavins, R. N. (1994). The energy paradox and the diffusion of conservation technology. *Resource and Energy Economics*, 16(2):91–122.

Kasy, M. and Sautmann, A. (2021). Adaptive treatment assignment in experiments for policy choice. *Econometrica*, 89(1):113–132.

Kitagawa, T. and Tetenov, A. (2018). Who should be treated? Empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2):591–616.

Knittel, C. R. and Stolper, S. (2019). Using machine learning to target treatment: The case of household energy use. Working Paper.

Manski, C. F. (2004). Statistical treatment rules for heterogeneous populations. *Econometrica*, 72(4):1221–1246.

Manski, C. F. (2021). Econometrics for decision making: Building foundations sketched by Haavelmo and Wald. *Econometrica*, 89(6):2827–2853.

Myers, E. and Souza, M. (2020). Social comparison nudges without monetary incentives: Evidence from home energy reports. *Journal of Environmental Economics and Management*, 101:102315.

Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. (2008). Normative social influence is underdetected. *Personality and Social Psychology Bulletin*, 34(7):913–923.

Prest, B. C. (2020). Peaking interest: How awareness drives the effectiveness of time-of-use electricity pricing. *Journal of the Association of Environmental and Resource Economists*, 7(1):103–143.

Savage, L. J. (1951). The theory of statistical decision. *Journal of the American Statistical association*, 46(253):55–67.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18(5):429–434.

Stoye, J. (2009). Minimax regret treatment choice with finite samples. *Journal of Econometrics*, 151(1):70–81.

Wald, A. (1950). *Statistical decision functions*. Wiley.

Appendix A Data Appendix

There are 23,829,510 observations of electricity meter readings in the original data. We remove instances of duplicate readings and reading periods of more than 35 days or less than 24 days. The top percentile of monthly usage is also removed to eliminate instances of implausibly high consumption levels. These sample restrictions reduce the sample to 21,921,882 observations.

As discussed in Section 3.2, the estimation sample consists of waves three, six, and seven due to data limitations for all other waves. The second and third columns of Table A.1 summarize the number of observations and number of unique accounts for these waves. Because our targeting analysis requires data on additional household characteristics, we further restrict the sample to households with non-missing and non-zero values for: 12-month average pre-treatment consumption, 12-month average post-treatment consumption, income, house size, building size, number of household members, and house age. The size of the final estimation sample for the pooled regression analysis is summarized in columns 4 and 5 of Table A.1. The pooled targeting analysis utilizes the cross-sectional version of this estimation sample, summarized in column 5.

Table A.1: Estimation sample

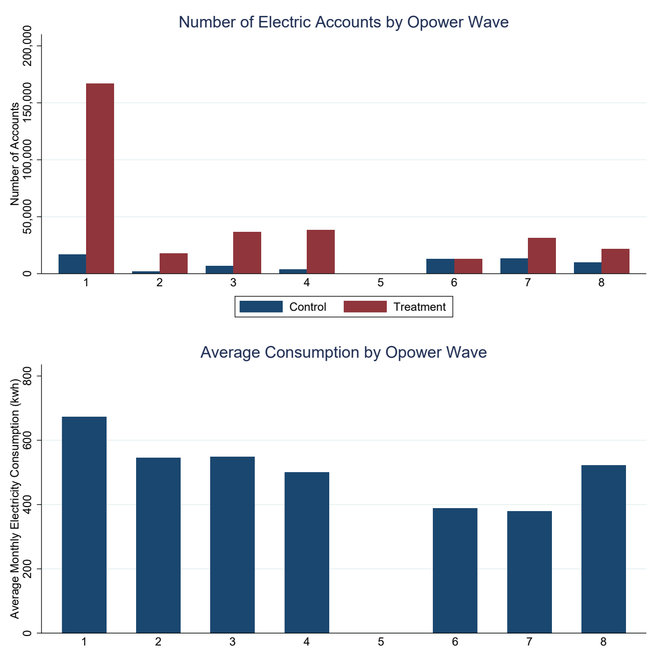
	Number of observations	Number of unique accounts	Number of observations with complete covariates	Number of unique accounts with complete covariates
Wave 3	2,089,684	43,435	1,167,703	22,915
Wave 6	1,126,998	25,974	504,889	11,114
Wave 7	1,324,613	44,372	513,513	15,507
Total	4,541,295	113,781	2,186,105	49,536

Notes: The number of unique accounts in column 3 include all households for wave 3, 6, and 7 after initial data cleaning. The number of unique accounts in column 5 include households with non-missing and non-zero values for covariates for the three waves.

Appendix B Verification of Opower Random Assignment

This section presents evidence on the success of randomization of Opower treatment assignment and justifies the variations used for identification. As mentioned previously, sufficient pre- and post-treatment consumption data is only available for waves 3, 6, and 7, so we restrict attention to those three waves. The bottom panel of Figure B.1 shows a clear decreasing trend in electricity consumption across waves 3, 6, and 7. Households that consume relatively large amounts of electricity were enrolled into the program in earlier waves while later waves targeted lower usage households.

Figure B.1: Number of unique households and average monthly electricity consumption by wave



Notes: **Top**: Number of unique accounts in each of the eight Opower waves. **Bottom**: Average monthly electricity consumption for each Opower wave (before and after treatment). Consumption data available from Jan 2014 to April 2018.

Because of this selection into waves, we assess randomization by comparing treated and control units within each wave. Table B.1 compares the mean of covariates and pre-treatment consumption between the Opower control and treatment groups for wave 3. The balance test for the other two waves can be found in the Tables B.2 and B.3. There are no statistically significant differences in pre-treatment consumption or household characteristics between the Opower control and treatment groups in any of the waves.

Table B.1: Balance test for electric Opower wave 3

	Control	Treatment	Difference	t-statistic
12-month pre-treatment consumption (kWh)	650	647	3.56	0.46
Income ($)	72,487	72,786	-299	-0.37
Number of household members	2.56	2.55	.00624	0.20
Building size (ft^2)	3,681	3,744	-63.6	-0.51
Unit size (ft^2)	1,761	1,778	-17.1	-1.04
House Year Built	1951	1951	.025	0.04
Married	.575	.565	.0108	1.19

Table B.2: Balance of covariates for Opower wave 6

	Control	Treatment	Difference	t-statistic
12-month pre-treatment consumption (kWh)	451	445	6.29	1.02
Income ($)	64,085	64,641	-556	-0.69
Number of household members	2.02	1.98	.032	1.16
Building size (ft^2)	5,439	5,371	67.6	0.29
Unit size (ft^2)	1,794	1,793	.536	0.03
House Year Built	1948	1949	-.983	-1.43
Married	.432	.428	.00414	0.44

Table B.3: Balance of covariates for Opower wave 7

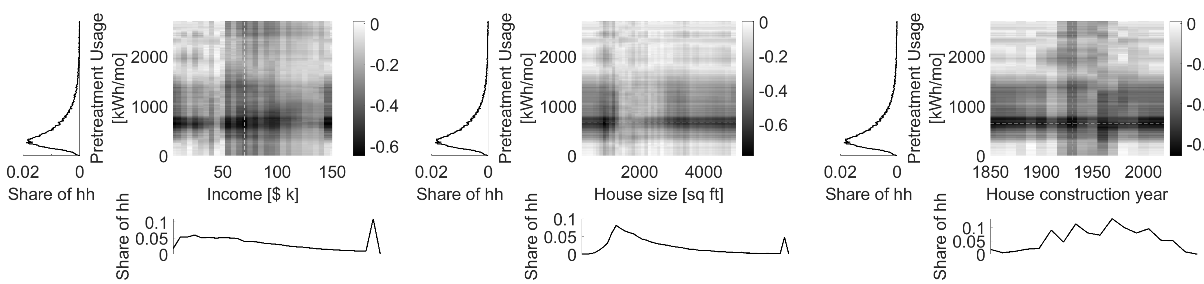
	Control	Treatment	Difference	t-statistic
12-month pre-treatment consumption (kWh)	479	487	-7.66	-1.13
Income ($)	53,653	54,389	-735	-1.02
Number of household members	1.88	1.89	-.0157	-0.65
Building size (ft^2)	5,660	5,778	-118	-0.43
Unit size (ft^2)	2,186	2,201	-14.6	-0.70
House Year Built	1937	1937	.539	0.85
Married	.322	.33	-.00736	-0.90

Appendix C Computational Details of EWM Estimation

This section describes the computational process solving for the EWM rules. We use grid search to find the optimal quadrant rule and IBM's CPLEX Optimizer to solve for the optimal linear rule with cubic terms.

Grid Search As discussed in Section 5, we consider three pairs of characteristics for targeting: (1) pre-treatment electricity usage and income; (2) pre-treatment usage and house size; and (3) pre-treatment usage and house age. We discretize these characteristic spaces into uniform grids with column grid size as 10 kWh for baseline consumption and row grid size as \$5,000, 100 ft^2, and 10 years for income, house size, and house age respectively. At each grid point, we calculate the net program impact of each of the four candidate quadrant rules (one for each direction) and pick the quadrant rule that yields the highest benefit. This step is repeated for each grid point in the characteristic space in order to identify the grid point with the highest net benefit. Figure C.1 plots the savings associated with the best quadrant rule (among the four candidate quadrants) for each grid point as a heatmap.

Figure C.1: EWM quadrant rule grid search heatmap



Notes: Heatmaps illustrate the private cost reduction per household per month estimated using the EWM method, with negative values representing effective cost savings. The solution of the EWM quadrant rule is the quadrant of the grid node with the most aggregate cost savings (darkest black). Each heatmap is complemented by density plots of the marginal distributions of each characteristic.

Linear Programming The optimal linear rule with cubic terms is represented by five parameters $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ in equation 8. As explained in the empirical strategy section, the linear rule with cubic terms is the solution to the mixed-integer linear programming problem specified in equation 10, reproduced here for convenience:

$$\min_{z,\beta} \frac{1}{N} \sum_{i=1}^{N} g_i \, z_i \quad \text{s.t.} \quad \frac{X_i^T \beta}{c_i} < z_i \leq 1 + \frac{X_i^T \beta}{c_i} \; \forall i, \quad z_i \in \{0, 1\}$$

To solve this problem using IBM's CPLEX Optimizer, we rewrite it as follows:

$$\min_{h} \quad f^T h \quad \text{s.t.} \quad Ah \leq b$$

where

$$f = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ g_1 \\ \vdots \\ \vdots \\ g_n \end{bmatrix}, \; h = \begin{bmatrix} \beta_0 \\ \vdots \\ \vdots \\ \beta_4 \\ z_1 \\ \vdots \\ \vdots \\ z_n \end{bmatrix}, \; A = \begin{bmatrix} 1 & -x1_1 & -x1_1^2 & -x1_1^3 & -x2_1 & c_1 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \ddots & \vdots & \vdots & \vdots \\ 1 & -x1_n & -x1_n^2 & -x1_n^3 & -x2_n & 0 & \dots & \dots & 0 & c_n \\ 1 & x1_1 & x1_1^2 & x1_1^3 & x2_1 & -c_1 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \ddots & \vdots & \vdots & \vdots \\ 1 & x1_n & x1_n^2 & x1_n^3 & x2_n & 0 & \dots & \dots & 0 & -c_n \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & 0 & 1 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & -1 \end{bmatrix}, \text{ and } b = \begin{bmatrix} c_1 \\ \vdots \\ c_n \\ 0_{n,1} \\ 0_{n-1,1} \end{bmatrix}.$$

For computational efficiency, households are aggregated into n pseudo-households with identical covariate values in the two-dimensional characteristic space, which we refer to as "unique households." We solve for β at the unique households level and then use $X^T\beta$ to derive the treatment rule at the households level.

To reduce the computational burden of the optimization procedure, we discretize the two-dimensional characteristic space into 30 equal bins along each dimension. The resulting bins have widths of 91 kWh/mo for pre-treatment consumption, 167 ft^2 for house size, and 5.5 years for house age. The outcome variable and covariates are normalized to the $[-1, 1]$ range by dividing by the absolute maximum value of the demeaned value for computational purposes.

Details of Kitagawa & Tetenov's Confidence Intervals The two-sided confidence interval for $V(\widehat{\pi})$ is

$$\left[V_n(\widehat{\pi}) - \frac{\hat{q}_{\tilde{v}}(1-\bar{\alpha})}{\sqrt{n}} \quad , \quad V_n(\widehat{\pi}) + \frac{\hat{q}_{\tilde{v}}(1-\bar{\alpha})}{\sqrt{n}} \right]$$

where $V_n(\widehat{\pi})$ is the savings point estimate from the original sample (denoted by the subscript n). To ensure adequate coverage, $\tilde{v}_n$ is specified as $\sqrt{n}\sup_{\pi\in\Pi}|V_n(\pi) - V(\pi)|$. Confidence intervals are constructed via bootstrap by resampling households with replacement 1,000 times for quadrant rules and 300 times for cubic rules, computing the maximum of the absolute value of the difference between savings in the bootstrap sample and savings in the original sample across all candidate rules (within each bootstrap sample), and then using the 95th percentile of the distribution across bootstrap samples for $\hat{q}_{\tilde{v}}(1-\bar{\alpha})$. See Kitagawa and Tetenov (2018) Appendix B for details on the asymptotic properties of this approach.

Comparison of *Specific* Estimated EWM Rules and the RCT We formally test whether the EWM rules estimated on the original sample outperform the actual RCT assignment. Let θ be the net savings from a given EWM rule relative to the RCT assignment in the population. We estimate this by taking the difference between the two savings point estimates in the original sample, both of which are relative to a baseline of no treatment. This is $\hat{\theta}$. To conduct inference, we resample households with replacement 1,000 times and calculate the difference between the savings from applying the *original* EWM rule to each bootstrap sample and the scaled average treatment effect on the treated in each bootstrap sample. This yields a distribution of net savings estimates across bootstrap samples. Confidence intervals are calculated as $\left[\hat{\theta} - z_{\alpha/2}\widehat{SE}\ ,\ \hat{\theta} + z_{\alpha/2}\widehat{SE}\right]$, where $z_{\alpha/2}$ is the critical value of the standard normal distribution (with $\alpha = 0.05$) and $\widehat{SE}$ is the standard error of $\hat{\theta}$ across bootstrap samples.

Appendix D Out-of-Sample Performance

D.1 Targeting using Historical Data: In- vs Out-of-Sample Performance

The main analysis in Section 6 use the entire pooled sample to estimate rules and to evaluate their performance. The results serve as a proof of concept. However, using the entire pooled sample to evaluate rules' performance does not provide information on out-of-sample fit, or whether the approach we use is susceptible to overfitting. One way we address this possibility is by using past waves to derive treatment rules for future waves in Section 6.5. We then evaluate their performance based on the future waves rather than the past waves. This provides some insight into out-of-sample performance.

To facilitate comparisons between in-sample and out-of-sample performance when targeting using historical data, we summarize the in-sample performance in Table D.1. When the sample wave used to estimate the rules is also used to quantify energy and cost savings, we find that the rules consistently outperform randomized treatment assignment under the original experiment. This is consistent with the qualitative results for the entire pooled sample in Section 6.

Table D.1: In-sample estimates of the gains from targeting using historical data

Target wave	Sample wave	Pre-treatment characteristics used for targeting	Energy changes kWh/hh-month	Private cost changes $/hh-month	Social cost changes $/hh-month
6	3	Income and mean usage	-7.22	-1.40	-0.80
		House size and mean usage	-12.95	-2.41	-1.03
		House age and mean usage	-10.6	-2.10	-0.98
7	6	Income and mean usage	-7.73	-1.30	-0.58
		House size and mean usage	-7.49	-1.36	-0.68
		House age and mean usage	-7.55	-1.38	-0.61

Notes: Net energy and cost changes are calculated by applying the EWM rule estimated on the re-weighted sample wave to re-weighted data from the sample wave. Negative values indicate reductions in energy consumption and cost, all of which are relative to the outcomes attained under the program's actual treatment assignment. All EWM rules are linear rules with cubic terms.

Table D.2 reproduces the out-of-sample performance presented in Table 5. This table evaluates performance for the same targeted treatment assignment rules that are evaluated in Table D.1, but it uses the later target wave instead of the earlier sample wave to do so. The estimates can be interpreted as the *ex-post* savings that would have been achieved had the treatment rule estimated on the earlier wave been applied to the later wave in practice. These *ex-post* savings are smaller in magnitude than their counterparts in Table D.1. This could be due to unobserved differences in households across the waves, overfitting, or a combination of the two. We explore the possibility of overfitting in additional analysis below. However, even though prospective targeting using historical data is not as effective as targeting *ex-post*, the results show that it is possible to outperform randomization in practice.

Table D.2: Out-of-sample estimates of the gains from targeting using historical data

Target wave	Sample wave	Pre-treatment characteristics used for targeting	Energy changes kWh/hh-month	Private cost changes $/hh-month	Social cost changes $/hh-month
6	3	Income and mean usage	-0.50	-0.06	-0.22
		House size and mean usage	0.08	0.13	-0.01
		House age and mean usage	-1.80	-0.33	-0.23
7	6	Income and 21mean usage	-0.17	0.00	-0.41
		House size and mean usage	-2.42	-0.11	-0.39
		House age and mean usage	1.37	-0.08	-0.41

Notes: This is a reproduction of Table 5 from the main text. Net energy and cost changes are calculated by applying the EWM rule estimated on the sample wave to data from the target wave. These results are point estimates of the *ex-post* savings that would have been achieved had the treatment rule estimated on the earlier wave been applied to the later wave in practice. Negative values indicate reductions in energy consumption and cost, all of which are relative to the outcomes attained under the program's actual treatment assignment. All EWM rules are linear rules with cubic terms.

D.2 Pooled Sample: Out-of-Sample Performance

One shortcoming of the approach above is that differences in the distribution of household characteristics across waves could confound comparisons of in-sample and out-of-sample performance. To provide additional evidence on the performance of our approach, we randomly divide the pooled sample into observably similar training and testing data and use them to evaluate in-sample and out-of-sample performance.

First, we randomly split the pooled sample into a training set containing half of the households and a testing set containing the other households. We stratify by treatment status and wave in the original experiment to ensure comparable wave-specific treatment shares in the training and testing set. Next, we use the training set to learn targeted treatment assignment rules. We then use the testing set to estimate the savings associated with each of these rules. Finally, we repeat this procedure 100 times and average the results.

Since the training and testing sets are drawn from the same distribution, the estimated savings based on the testing set provide a measure of out-of-sample performance that is not confounded by differences across groups of households. This provides a way to diagnose overfitting: if the treatment assignments derived by the EWM method are highly sensitive to sampling variation in the training data, they would not perform well in the testing data.

Table D.3 shows the estimated energy savings, private cost savings, and social cost savings for targeted treatment rules evaluated using the testing data. We use the mean and standard deviation across the 100 training/testing splits to summarize out-of-sample performance. In almost all cases, the rules outperform a benchmark of no treatment. In cases where they do not, the cost changes are close to zero. The rules compare favorably to the original experiment for maximizing private and social cost savings. For social cost savings, all of the rules outperform randomized treatment assignment under the original experiment on average.

The targeted treatment assignment rules in Table D.3 vary in their flexibility and inputs. The one-dimensional rules based on average pre-treatment electricity consumption alone tend to outperform more flexible rules based on additional data. Similarly, quadrant rules exhibit better out-of-sample performance than cubic rules. These results suggest that more flexible rules are susceptible to overfitting. Overall, however, the EWM method outperforms randomized treatment assignment and no treatment assignment, both in and out of sample.

Table D.3: Out-of-sample estimates of the gains from targeting for the pooled sample

Treatment rule	Pre-treatment characteristics used for targeting	Energy changes kWh/hh-month	Private cost changes $/hh-month	Social cost changes $/hh-month
Actual RCT	–	-3.639 $\pm$ 1.87	-0.093 $\pm$ 0.33	0.315 $\pm$ 0.12
EWM-univariate	Mean usage	-3.502 $\pm$ 1.97	-0.397 $\pm$ 0.33	-0.013 $\pm$ 0.08
EWM-quadrant	Income and mean usage	-2.921 $\pm$ 1.66	-0.251 $\pm$ 0.29	-0.006 $\pm$ 0.09
EWM-quadrant	House size and mean usage	-3.796 $\pm$ 1.86	-0.356 $\pm$ 0.34	-0.029 $\pm$ 0.10
EWM-quadrant	House age and mean usage	-2.574 $\pm$ 1.71	-0.274 $\pm$ 0.31	-0.009 $\pm$ 0.09
EWM-cubic	Income and mean usage	-3.354 $\pm$ 1.83	-0.087 $\pm$ 0.29	-0.025 $\pm$ 0.10
EWM-cubic	House size and mean usage	-2.589 $\pm$ 1.89	0.030 $\pm$ 0.29	0.034 $\pm$ 0.09
EWM-cubic	House age and mean usage	-2.848 $\pm$ 1.80	-0.136 $\pm$ 0.32	0.008 $\pm$ 0.09

Notes: This table summarizes out-of-sample performance for EWM rules that are derived from training data. Energy and cost changes are computed using testing data. This procedure is repeated 100 times, and each cell of the table presents the mean and standard deviation of changes across repetitions. Negative values indicate reductions in energy consumption and cost, all of which are relative to no treatment.

Appendix E Targeting with Budget Constraints

In this appendix, we summarize the benefits of targeting in the presence of budget constraints. Imposing constraints reduces the flexibility of treatment assignment rules relative to an unconstrained approach, and therefore will not perform as well in general. However, there are multiple reasons why it may still be desirable to impose constraints. First, program administrators might face a budget cap on total program expenses, and therefore might need to design treatment rules that will satisfy that budget cap. For programs in which the marginal cost of treating each household is constant, a budget cap is equivalent to a cap on the share of households eligible to receive treatment. Second, there might be situations when the program administrator prefers to fix, rather than cap, a program's budget. This is equivalent to fixing the share of treated households. We analyze these two types of budget constraints in more detail below.

E.1 Targeting with a Budget Cap

Using the original program as the benchmark, we set the treatment cap equal to the share of households that actually received treatment in the pooled sample. Following the method proposed in Kitagawa and Tetenov (2018), we add a "rationing" term to the objective function. Under this approach, for any candidate rule that would treat a larger number of households than the original experiment, treatment would be rationed randomly to a subset of households assigned to treatment in order to satisfy the budget cap. The resulting objective function is

$$\widehat{\pi}^k \in \arg\min_{\pi\in\Pi} \left\{ \frac{1}{N}\sum_{i=1}^{N}\left[\left(\frac{Y_i D_i}{e(X_i)} - \frac{Y_i(1-D_i)}{1-e(X_i)}\right)\cdot \mathbb{1}_{(X_i\in\pi)}\right]\cdot \underbrace{\min\left\{1, \frac{K}{P_X(\pi)}\right\}}_{\text{rationing term}} \right\}, \qquad (12)$$

where the final term inside the larger brackets is the new rationing term. K denotes the treatment share cap, and $P_X(\pi)$ represents the share of households that would be treated under a given treatment rule π. For any candidate rule that would treat more than K households, the cap binds, so treatment would be randomly assigned to a fraction $K/P_X(\pi)$ of the households that would otherwise be treated under rule π absent a budget cap. The rationing term in equation 12 scales down the estimated benefits of targeting from the unconstrained rule to account for the effect of randomizing treatment in order to satisfy this constraint. For any candidate rule that would treat fewer households than the original experiment, the rationing term would be one, as no rationing is needed. In this case, the estimated benefits of targeting would be the same as before, and the budget cap would not bind. After making this modification, we repeat the grid search procedure described in Section 5 and Appendix C in order to search for the quadrant that minimizes expected electricity consumption and cost as defined in equation 12.[27]

Table E.1 summarizes the energy and cost savings that would be achieved by targeted treatment rules that satisfy a budget cap defined by the share of households treated in the original experiment. All savings estimates are relative to a baseline of no treatment. The table also includes the treatment share for each rule. The first row contains the results from the original experiment as a benchmark.

Table E.1: Estimates of the gains from targeting with a budget cap

Treatment Rule	Pre-treatment characteristics used for targeting	Energy usage		Private cost		Social cost	
		Share treated	Changes	Share treated	Changes	Share treated	Changes
		%	kWh/hh-mo	%	$/hh-mo	%	$/hh-mo
Actual RCT	–	72	-3.63	72	-0.09	72	0.32
EWM-univariate	Mean usage	31	-4.93	29	-0.64	5	-0.12
EWM-quadrant	Income and mean usage	31	-4.93	13	-0.65	13	-0.18
EWM-quadrant	House size and mean usage	28	-5.63	27	-0.79	22	-0.17
EWM-quadrant	House age and mean usage	67	-5.10	21	-0.66	4	-0.15

Notes: This table presents point estimates of the savings from targeted treatment assignment rules that are constrained to treat no more than the number of households treated by the original experiment. Negative changes indicate reductions in energy consumption and cost. All estimates are normalized to represent the savings per household in the sample, not savings per household treated, in order to facilitate comparisons across rules that treat different numbers of households.

All of the targeted treatment assignment rules in Table E.1 would treat fewer households, and achieve greater savings, than the original experiment.[28] For treatment rules designed to minimize energy consumption, the optimal rules presented in Table G.2 are no longer feasible, because they exceed the budget cap. This implies that the budget-constrained EWM rules should achieve lower energy savings than that of the unconstrained EWM rules. That is indeed what we find. However, the constrained rules still outperform the original experiment.[29] In contrast, the unconstrained EWM rules that maximize private and social marginal cost savings already meet the budget cap. For those rules, the savings estimates in Table E.1 are identical to the unconstrained case,

[27]We present results for the one-dimensional and quadrant rules. We did not reproduce these constrained analyses for cubic rules due to their added computational complexity.

[28]The energy and cost savings per household for each treatment rule and for the original experiment represent the average of all households in the sample, not just the households that would be treated by a specific rule. This is consistent with our presentation of results throughout the paper.

[29]Interestingly, using different rules that treat a smaller fraction of households generates greater savings than rationing treatment within the households treated by the unconstrained rules, which would meet the budget exactly.

summarized in Tables G.3 and G.4.

E.2 Targeting with a Fixed Budget

To study the case of a fixed budget, we further constrain the candidate rules to treat the same number of households as the original experiment. This rules out any treatment assignment rule that would treat fewer households. For rules that would treat more households than the original experiment, we follow the rationing method described above.

Table E.2 summarizes the energy and cost savings that would be achieved by targeted treatment rules that treat the same number of households as the original experiment. These rules continue to outperform the original experiment. However, the strict budget constraint reduces the potential effectiveness of targeting. In all cases, the savings estimates are lower in magnitude than the corresponding estimates from the budget cap analysis (Table E.1) and the unconstrained analysis in the main text (Tables G.2, G.3, and G.4). The point estimates for the social marginal cost analysis in Table E.2 are actually positive, which suggests that constraining the program to treat a fixed number of households may be worse than treating none at all. This highlights a potential downside of constraining programs to meet a fixed budget. Furthermore, it underscores that the promise of targeted treatment assignment comes not just from choosing which households to treat, but also how many.

Table E.2: Estimates of the gains from targeting with a fixed budget

Treatment Rule	Pre-treatment characteristics used for targeting	Energy usage		Private cost		Social cost	
		Share treated	Changes	Share treated	Changes	Share treated	Changes
		%	kWh/hh-mo	%	$/hh-mo	%	$/hh-mo
Actual RCT	–	72	-3.63	72	-0.09	72	0.32
EWM-univariate	Mean usage	72	-4.10	72	-0.18	72	0.28
EWM-quadrant	Income and mean usage	72	-4.53	72	-0.25	72	0.26
EWM-quadrant	House size and mean usage	72	-4.79	72	-0.30	72	0.24
EWM-quadrant	House age and mean usage	72	-5.06	72	-0.34	72	0.23

Notes: This table presents point estimates of the savings from targeted treatment assignment rules that are constrained to treat the same number of households as the original experiment. Negative changes indicate reductions in energy consumption and cost. All estimates are normalized to represent the savings per household in the sample, not savings per household treated, in order to facilitate comparisons with other rules that treat different numbers of households.

Appendix F Distributional Implications of Targeting

Based on the analysis in this paper, targeted treatment rules would produce significant improvements in program performance. The measures of performance we consider are total electricity consumption, private cost, and social cost. We chose the first measure because home energy report programs are intended to encourage energy conservation. The two cost measures are alternative measures that capture many – though not all – of the components of traditional economic measures of aggregate welfare impacts.

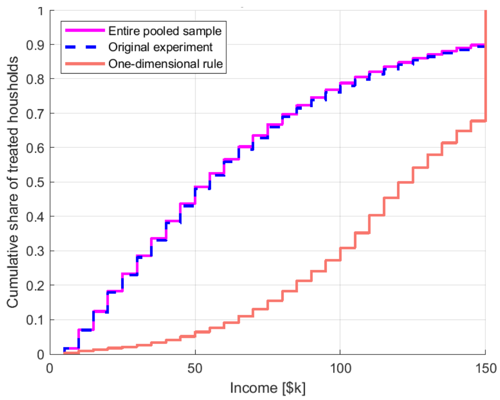
However, there may be distributional implications of targeted treatment assignment that our analysis does not account for. This is most evident in the case of treatment rules that explicitly condition on measures like household income. Even without using sensitive characteristics like income directly, targeting based on pre-treatment electricity consumption could yield outcomes that treat people of different incomes or racial groups differently. In this appendix, we document how treatment rules affect households of different incomes and racial groups, and we highlight how program design choices could influence these distributional implications.

F.1 Income

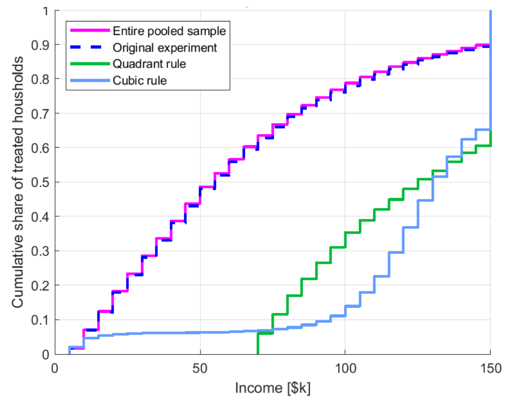
To summarize how different treatment assignment rules affect households with different incomes, we compare the empirical cumulative distribution function (CDF) of income for each rule. Figure F.1 summarizes the results. First, in Figure F.1a, we plot the CDF for three different groups: the entire pooled sample, the treated households within the pooled sample, and the households that would be treated by the one-dimensional EWM rule that uses mean pre-treatment electricity consumption to maximize private cost savings. The entire pooled sample and the subsample that was treated by the original experiment exhibit almost identical distributions of income. This is an artifact of randomization of treatment in the original experiment. In contrast, using a strategy to increase cost savings based on pre-treatment electricity consumption has the effect of treating high-income households at a disproportionately high rate, and low-income households at a disproportionately low rate. This is because households in the data with higher pre-treatment electricity consumption tend to have higher incomes.

More flexible rules yield different patterns. In some cases, income is an explicit input into the optimization problem. For those rules, the distribution of income is similar to the one-dimensional rule that is only based on pre-treatment electricity consumption, as can be seen in Figure F.1b. The one-dimensional rule in Figure F.1a and the cubic rule in Figure F.1b are strikingly similar. This comparison suggests that using pre-treatment electricity consumption alone effectively serves as a proxy for income due to their correlation in the data. The quadrant rule in Figure F.1b is even more extreme insofar as no households with incomes under $70,000 per year are treated.

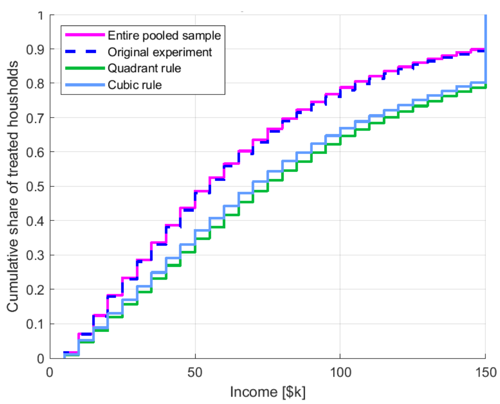
Figure F.1: Empirical CDF of income for treated households under various treatment assignment rules



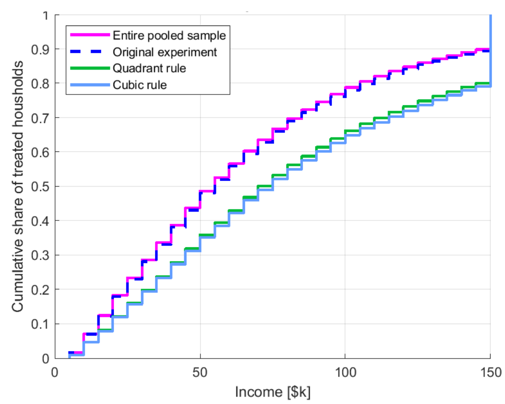
(a) One-dimensional EWM rule: mean usage



(b) EWM rules: income and mean usage



(c) EWM rules: house size and mean usage



(d) EWM rules: min and mean usage

Notes: These figures summarize the distributional implications of targeted treatment assignment across households of different incomes. The figures contain empirical cumulative distribution functions (CDFs) of income for the entire pooled sample (solid magenta), the treatment group in the original experiment (dashed blue), the treatment group of the one-dimensional EWM rule (solid red), treatment groups of the EWM quadrant rules (solid green), and treatment groups of the EWM cubic rules (solid blue). All EWM rules shown in the figure target treatment to minimize private energy cost.

In contrast, for rules that incorporate other demographic characteristics like house size, the treated sample is more similar to the entire pooled sample (Figure F.1c). These rules still treat high-income households at a higher rate than randomized treatment assignment, but the difference is modest. Finally, for rules that only use pre-treatment electricity consumption, but incorporate both the mean and the minimum of consumption, the CDFs are more similar to the entire pooled sample (Figure F.1d).[30] Including an

[30]Targeting rules based on mean pre-treatment electricity consumption in combination with house

additional measure of electricity consumption has the effect of undoing the correlation between average consumption and income that creates distributional impacts that may be undesirable. This suggests that a strategy of incorporating multiple measures of pre-treatment electricity consumption data may allow program administrators to achieve the benefits of targeting based on pre-treatment electricity consumption without creating significant distributional implications.

F.2 Race

Our data do not contain household-level information on race, so we are unable to construct direct measures of racial diversity for the treatment group under each rule as we did for income. However, we do observe the zip code each household is located in. Thus, we collect data on the racial composition of zip codes from the U.S. Census Bureau's 2015-2020 American Community Survey to serve as a proxy for the racial composition of households in the sample. We use these data to construct a measure of racial diversity based on the Diversity Index from the 2020 Census. The index is a number between zero and one that can be interpreted as the probability that two individuals selected at random from within a group are of different races. Higher levels of the index correspond to greater racial diversity.

We calculate the diversity index as $\sum_{r=1}^{R} s_r(1 - s_r)$, where s_r is the share of households in a group that are of race r.[31] We use the American Community Survey data to categorize residents into six single-race groups (White, Black, American Indian, Asian, Hawaiian or Pacific Islander, Some Other Race) and one group for those who identify as members of two or more races. We then compute the share of residents in each racial group (i.e., s_r) in each zip code, and use this to construct the diversity index for each zip code. Figure F.2 shows the distribution of the diversity index across zip codes.

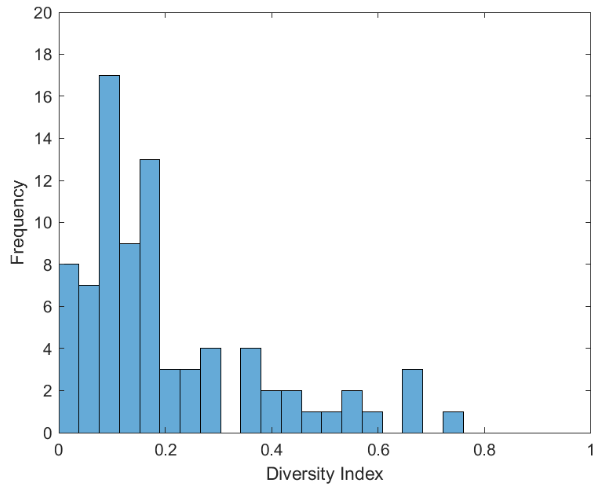
We use the diversity index scores to visualize the correlation between each zip code's diversity index and the share of households treated by each rule in that zip code. A positive correlation would indicate that treatment is assigned at a higher rate in more diverse zip codes. A negative correlation would indicate that treatment is assigned at a higher rate in more racially homogeneous zip codes.

Figure F.3 summarizes these correlations. We include data on treatment assignment in the original experiment in each panel to serve as a benchmark for the targeted treatment assignment rules. Figure F.3a visualizes the relationship between racial diversity

vintage, maximum electricity consumption, or the standard deviation of electricity consumption all yield similar income distributions. Thus, we omit them to save space.

[31]This is equivalent to one minus the Herfindahl-Hirschman index, a measure of market concentration.

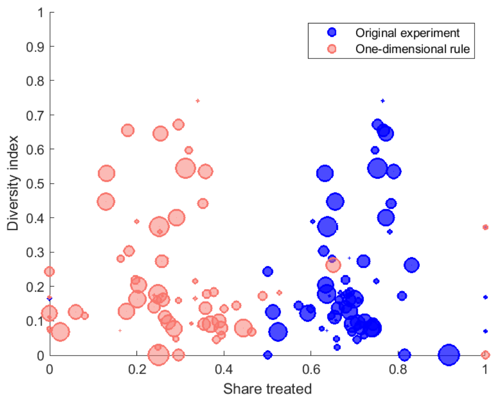
Figure F.2: Distribution of racial diversity index scores across zip codes



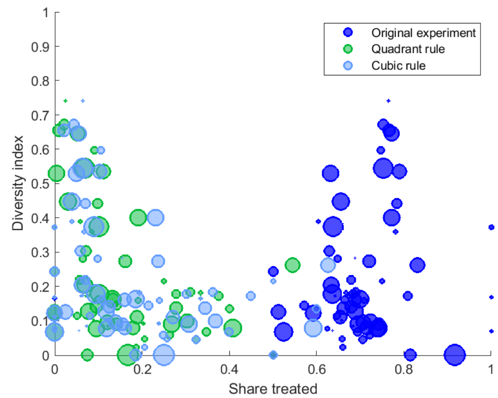
Notes: This figure shows the distribution of the diversity index based on data from the American Community Survey. The diversity index is calculated at the zip code level. Higher values indicate more racial diversity.

and treatment shares for the one-dimensional rule that uses pre-treatment electricity consumption and seeks to maximize private cost savings. The level of the treatment shares for the EWM rule are lower on average and somewhat more dispersed than for the original experiment. However, there is no clear correlation between racial diversity and treatment shares for either the original experiment or the targeted treatment assignment rule we derive. We find similar patterns for EWM rules that maximize private cost savings derived using both pre-treatment electricity usage and income, house size, or minimum pre-treatment usage (Figures F.3b, F.3c, and F.3d). Thus, while we cannot directly observe the race of each household assigned to treatment, we do not find evidence that using targeting to assign treatment introduces racial bias at the zip code level.

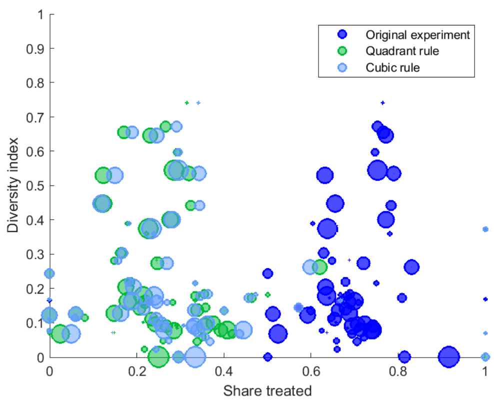
Figure F.3: Racial diversity index vs treatment share across zip codes under various EWM rules



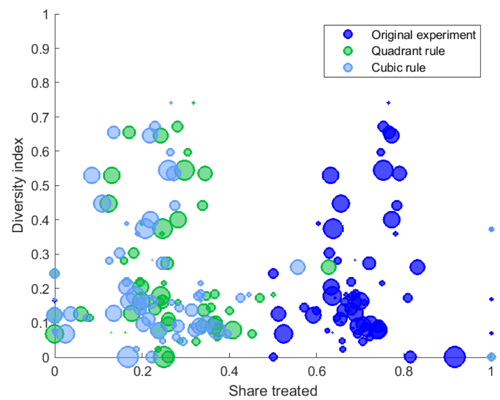
(a) One-dimensional EWM rule: mean usage



(b) EWM rules: income and mean usage



(c) EWM rules: house size and mean usage

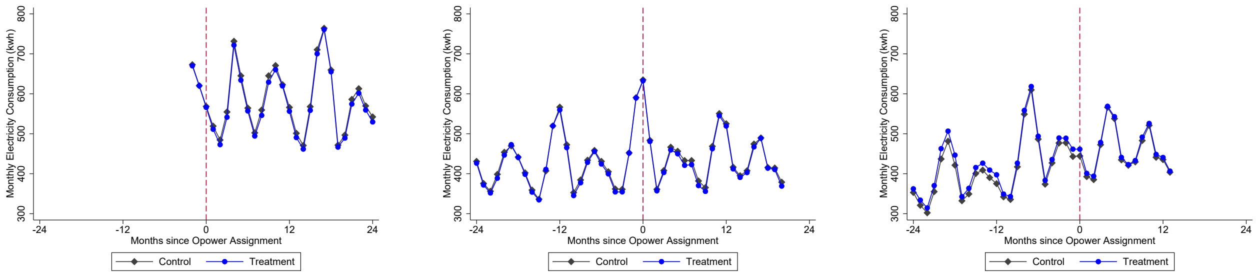


(d) EWM rules: min and mean usage

Notes: These figures present scatter plots of the share of households treated in each zip code on the x-axis and the diversity index of that zip code on the y-axis. Each figure compares the treatment share under the original experiment, in dark blue, to an alternative rule based on the EWM method. Each circle represents a zip code and the size of each circle is proportional to the number of treated households in that zip code. All EWM rules shown in the figure target treatment to minimize private energy cost.

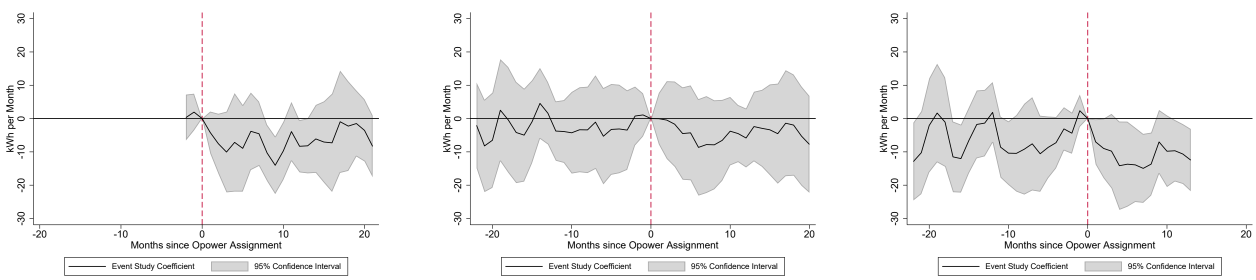
Appendix G Additional Figures and Tables

Figure G.1: Average electricity consumption by treatment arm



Notes: Average electricity consumption by treatment arm around the time of treatment for wave 3 (**left**), 6 (**middle**), and 7 (**right**). Each point represents the mean electricity consumption in a given event month and treatment arm. Event time is normalized to zero in the first month of treatment.

Figure G.2: Event studies by wave



Notes: Treatment effect estimates by event month, estimated separately for wave 3 (**left**), 6 (**middle**), and 7 (**right**). The month of Opower assignment (0) is normalized to zero. Observations with event month prior to -22 are grouped to event month -23 and those with event month post event 21 are grouped to event month 22, and these endpoints are omitted from the plots. The endpoints are chosen based on the event months for wave 6, which was initiated close to the middle of sample. A time-invariant Opower treatment indicator, household characteristics, and wave-by-year-month fixed effects are included in the event study model but omitted from these plots.

Table G.1: Robustness of average treatment effect estimates

	Dependent variable: Electricity Usage in kWh			
	(1)	(2)	(3)	(4)
	All waves	Wave 3	Wave 6	Wave 7
Opower × Post - both	-5.79**	-6.31	-0.96	-3.11
	(2.55)	(4.83)	(2.49)	(2.69)
Opower × Post - no demographics	-5.47**	-6.26	-0.95	-3.00
	(2.62)	(4.83)	(2.50)	(2.69)
Opower × Post - no baseline usage	-9.52**	-6.43	-1.89	-8.20**
	(4.11)	(4.91)	(2.94)	(4.12)
Opower × Post - neither	-8.91**	-6.39	-1.43	-6.45
	(4.35)	(4.91)	(3.00)	(4.32)
Wave × year-month FE	Yes	Yes	Yes	Yes
Control mean	473	579	428	422
Households	49,536	22,915	11,114	15,507
Observations	2,186,105	1,167,703	504,889	513,513

* p<0.10, ** p<0.05, *** p<0.01

Notes: This table shows average treatment effect estimates from alternative specifications of equation 1. The first row shows the estimated coefficients including both demographics and baseline usage, as in Table 3. The second row shows the estimated coefficients omitting the demographics control variables. The third row shows the estimated coefficients omitting the baseline usage. The fourth row shows the estimated coefficients omitting both the demographics control variables and the baseline consumption. Wave-by-year-month FE indicates fixed effects for the interaction of wave and calendar sample month. Control means are the mean monthly electricity consumption for the control group in each estimation sample. Standard errors are clustered at the household level and shown in parentheses.

Table G.2: Electricity usage reductions from the EWM method

Treatment rule	Pre-treatment characteristics used for targeting	Share treated %	Net energy changes kWh/ hh-month	Δ EWM v. RCT kWh/ hh-month
Actual RCT	–	72	-3.63	
			(-10.48, 0.41)	
EWM-quadrant	Income and mean usage	92	-5.50	-1.88
			(-11.36, 0.35)	(-4.20, 0.45)
EWM-quadrant	Size and mean usage	92	-5.93	-2.30
			(-11.86, 0.00)	(-4.66, 0.06)
EWM-quadrant	House age and mean usage	78	-5.46	-1.83
			(-11.43, 0.51)	(-4.38, 0.72)
EWM-cubic	Income and mean usage	69	-7.43	-3.80
			(-14.49, -0.37)	(-5.89, -1.71)
EWM-cubic	Size and mean usage	92	-6.83	-3.20
			(-13.77, 0.11)	(-5.23, -1.16)
EWM-cubic	House age and mean usage	98	-6.60	-2.97
			(-13.43, 0.23)	(-5.02, -0.93)

Notes: This table shows estimates of the energy savings from the original experiment as implemented and the energy savings that would be achieved using EWM rules. The first row contains our estimate of the effect of the original experiment. All other rows correspond to EWM rules. Negative values indicate energy-usage reductions. The impact of the actual RCT is computed by multiplying the average treatment effect on the treated from the original experiment by the share of households treated to ensure a fair comparison between the RCT and EWM rules. The third column shows the share of households treated under each rule. The fourth column contains the estimated energy savings from applying the original random assignment and the derived EWM rules on the original sample, relative to no treatment. Two-sided 95% confidence intervals constructed using the method proposed in Kitagawa and Tetenov (2018) are in parentheses in column 4. The final column estimates the difference between the *specific* EWM rule estimated on the original sample and the actual RCT shown in column 4, with two-sided 95% confidence intervals constructed via bootstrap.

Table G.3: Net private cost reductions from the EWM method

Treatment rule	Pre-treatment characteristics used for targeting	Share treated %	Net cost changes $/ hh-month	Δ EWM v. RCT $/ hh-month
Actual RCT	–	72	-0.09	
			(-1.09, 0.84)	
EWM-quadrant	Income and mean usage	13	-0.65	-0.56
			(-1.68, 0.38)	(-1.1, -0.02)
EWM-quadrant	Size and mean usage	27	-0.79	-0.70
			(-1.84, 0.26)	(-1.05, -0.34)
EWM-quadrant	House age and mean usage	21	-0.66	-0.57
			(-1.72, 0.40)	(-1.02, -0.12)
EWM-cubic	Income and mean usage	15	-0.81	-0.72
			(-2.05, 0.44)	(-1.23, -0.21)
EWM-cubic	Size and mean usage	28	-0.76	-0.67
			(-1.97, 0.46)	(-0.98, -0.35)
EWM-cubic	House age and mean usage	26	-0.82	-0.73
			(-2.03, 0.39)	(-1.07, -0.39)

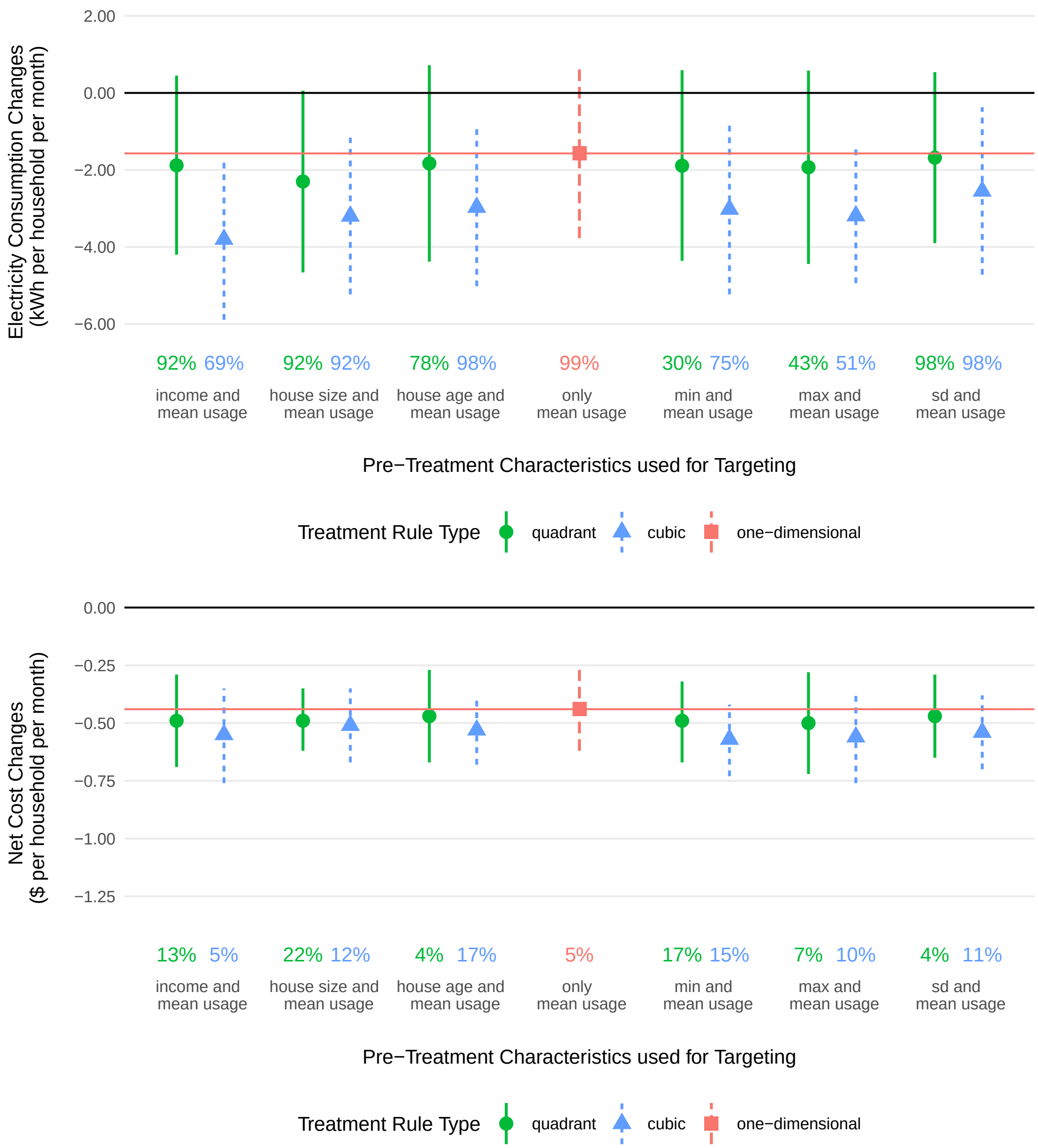
Notes: This table shows estimates of the private cost savings from the original experiment as implemented and the private cost savings that would be achieved using EWM rules. Net changes in cost are calculated as the the sum of electricity price ($0.177/kWh)*kWh reduction + program cost per household ($0.765/month) if treated. Negative values indicate reductions in cost. The impact of the actual RCT is computed by multiplying the average treatment effect on the treated from the original experiment by the share of households treated to ensure a fair comparison between the RCT and EWM rules. The third column shows the share of households treated under each rule. The fourth column contains the estimated cost savings from applying the original random assignment and the derived EWM rules on the original sample, relative to no treatment. Two-sided 95% confidence intervals constructed using the method proposed in Kitagawa and Tetenov (2018) are in parentheses in column 4. The final column estimates the difference between the *specific* EWM rule estimated on the original sample and the actual RCT shown in column 4, with two-sided 95% confidence intervals constructed via bootstrap.

Table G.4: Net social cost reductions from the EWM method

Treatment rule	Pre-treatment characteristics used for targeting	Share treated %	Net cost changes $/ hh-month	Δ EWM v. RCT $/ hh-month
Actual RCT	–	72	0.32	
			(0.08, 0.80)	
EWM-quadrant	Income and mean usage	13	-0.18	-0.49
			(-0.55, 0.20)	(-0.69, -0.29)
EWM-quadrant	Size and mean usage	22	-0.17	-0.49
			(-0.56, 0.22)	(-0.62, -0.35)
EWM-quadrant	House age and mean usage	4	-0.15	-0.47
			(-0.54, 0.23)	(-0.67, -0.27)
EWM-cubic	Income and mean usage	5	-0.24	-0.55
			(-0.70, 0.22)	(-0.76, -0.35)
EWM-cubic	Size and mean usage	12	-0.19	-0.51
			(-0.64, 0.25)	(-0.67, -0.35)
EWM-cubic	House age and mean usage	17	-0.22	-0.53
			(-0.65, 0.22)	(-0.68, -0.38)

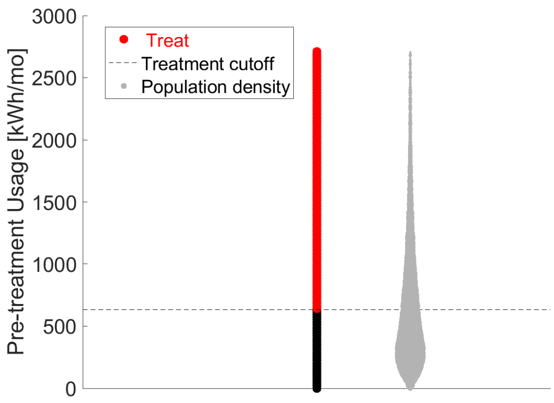
Notes: This table shows estimates of the social cost savings from the original experiment as implemented and the social cost savings that would be achieved using EWM rules. Net changes in social cost calculated as the the sum of Social Marginal Cost (SMC) ($0.065/kWh)*kWh reduction+program cost per household ($0.765/month) if treated. Negative values indicate reductions in cost. The impact of the actual RCT is computed by multiplying the average treatment effect on the treated from the original experiment by the share of households treated to ensure a fair comparison between the RCT and EWM rules. The third column shows the share of households treated under each rule. The fourth column contains the estimated cost savings from applying the original random assignment and the derived EWM rules on the original sample, relative to no treatment. Two-sided 95% confidence intervals constructed using the method proposed in Kitagawa and Tetenov (2018) are in parentheses in column 4. The final column estimates the difference between the *specific* EWM rule estimated on the original sample and the actual RCT shown in column 4, with two-sided 95% confidence intervals constructed via bootstrap.

Figure G.3: Comparison of the gains from targeting with and without using demographic characteristics

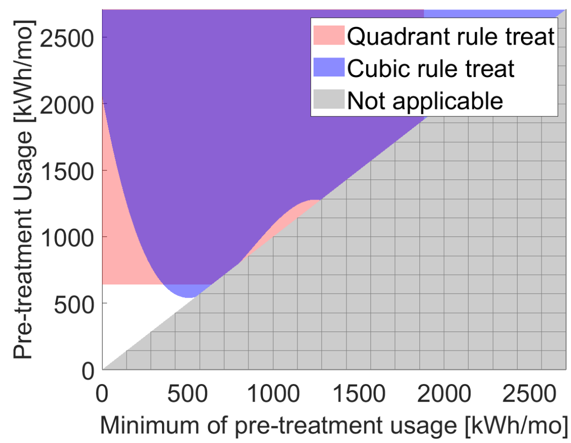


Notes: Points indicate estimates of the **energy savings** (top) and **social cost savings** (bottom) that would be achieved from targeted treatment assignment, relative to the original experiment as implemented. Lines represent 95% confidence intervals. Quadrant rules are shown in green solid lines, cubic rules are shown in blue dashed lines, and the one-dimensional rule is shown in red with long dashes. The percentage beneath each estimate represents the share of households that would be treated by that rule. Within each plot, the left side shows EWM rules based on both pre-treatment electricity consumption and demographic characteristics. The right side shows EWM rules based on pre-treatment electricity consumption only. To facilitate comparisons, a horizontal line indicates the savings associated with the one-dimensional EWM rules that use only average pre-treatment electricity consumption for targeting.

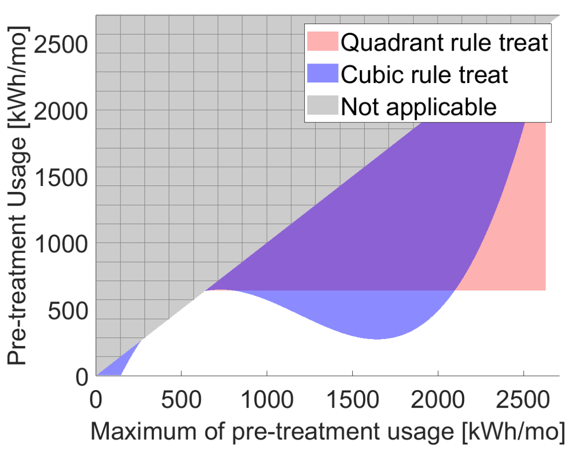
Figure G.4: EWM rules for private cost minimization using only pre-treatment consumption data



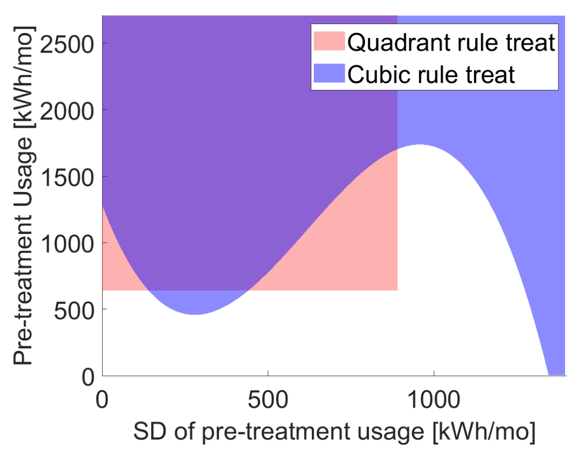
(a) average baseline consumption only



(b) average and minimum baseline consumption



(c) average and maximum baseline consumption



(d) average and standard deviation of baseline consumption

Notes: EWM rules based on functions of baseline consumption, where the objective is **private cost** minimization. Quadrant and cubic rules are indicated in red and blue shade in panel (b), (c) and (d).