Kyoto University,
Graduate School of Economics
Discussion Paper Series

Heterogeneous Treatment Effects of Nudge and Rebate:

Causal Machine Learning in a Field Experiment on Electricity Conservation

Kayo Murakami, Hideki Shimada, Yoshiaki Ushifusa, Takanori Ida

Discussion Paper No. E-20-003

Graduate School of Economics
Kyoto University
Yoshida-Hommachi, Sakyo-ku
Kyoto City, 606-8501, Japan

September, 2020

Heterogeneous Treatment Effects of Nudge and Rebate:

Causal Machine Learning in a Field Experiment on Electricity Conservation

September 2020

Kayo Murakami[*] **Hideki Shimada**[†] **Yoshiaki Ushifusa**[‡] **Takanori Ida**[§]

Abstract: This study investigates the different impacts of monetary and nonmonetary incentives on energy-saving behaviors using a field experiment conducted in Japan. We find that the average reduction in electricity consumption from rebate is 4%, while that from nudge is not significantly different from zero. Applying a novel machine learning method for causal inference (causal forest) to estimate heterogeneous treatment effects at the household level, we demonstrate that the nudge intervention's treatment effects generate greater heterogeneity among households. These findings suggest that selective targeting for treatment increases the policy efficiency of monetary and nonmonetary interventions.

JEL: D9, C93, Q4

Keywords: Causal Forest, Rebate, Nudge, Randomized Controlled Trial, Energy, Machine Learning

---

[*] Division of Natural Resource Economics, Graduate School of Agriculture, Kyoto University, Japan (murakamikayo@gmail.com), JSPS Post-Doctoral Research Fellow（RPD）(first author & 2nd corresponding author)

[†] Division of Natural Resource Economics, Graduate School of Agriculture, Kyoto University, Japan (hshimada.hs3@gmail.com), JSPS Research Fellow（DC2）(2nd first author)

[‡] Faculty of Economics and Business Administration, The University of Kitakyushu, Japan (ushifusa@kitakyu-u.ac.jp)

[§] Graduate School of Economics, Kyoto University, Japan (ida@econ.kyoto-u.ac.jp) (corresponding author)

1. Introduction

Incentive design is the central issue faced by public policymakers in the field of environmental conservation and resource management. In economic theory, negative externalities can be internalized by imposing Pigouvian taxes and subsidies. However, this may lead to typical decision-making obstacles related to pricing rules and financial resources in subsidies, which require substantial effort and time to overcome. In contrast, nonmonetary incentives, such as a nudge, are considered an alternative public policy tool due to their low implementation cost and ability to preserve freedom of choice. This study investigates the heterogeneous treatment effects of monetary and nonmonetary interventions on energy saving and explores a novel policy mix that could effectively utilize such heterogeneities.

Ample empirical evidence exists regarding the effects of monetary incentives in the context of residential energy conservation (Borenstein, 2002, 2005; Wolak, 2011; Joskow, 2012). For example, the treatment effect of critical peak pricing (CPP), which increases prices during critical peak hours, ranges between 7% and 22% peak-cut (Faruqui and Sergici, 2010; Jessoe and Rapson, 2014; Ito, Ida, and Tanaka, 2018). However, customer opposition often prevents the introduction of such pricing, as it could make some customers considerably worse off, making it difficult to implement as a public policy.[1] Although the treatment effect of equivalent amounts of rebate (the reward for energy conservation) is expected to be half or less than that of pricing (Wolak, 2011), reward-style incentives achieve regulatory approval more easily due to easier opt-in, which does not require a change in the basic electricity rate and encounters less consumer opposition.

Nonmonetary incentives refer to real-time feedback on the quantity of electricity consumed via an in-home display (IHD), moral suasion, and social comparison information on customers' energy conservation behaviors.[2] Using a field experiment, Jessoe and Rapson (2014) find that households with an IHD are more responsive to CPP.[3]

---

[1] The latest study based on a randomized controlled trial (RCT) field experiment (Fowlie et al., 2020) shows that the share of customers opting in for new pricing programs such as CPP and TOU (Time-of-Use tariff) is only approximately 20%, obviously lower than the 50% participation rate observed when all customers are informed and guaranteed positive gains from switching to a new pricing program (Ito, Ida, and Tanaka, 2017). The study's findings indicate that pricing-style incentives are hard to introduce in practice.

[2] In addition to these nonmonetary incentives, Azarova et al. (2020) recently demonstrated that interventions based on altruistic motives and collective action flaming affect customers' electricity reduction. Additionally, Arimura et al. (2016) investigate peer effects on energy consumption.

[3] Martin and Rivers (2018) present a detailed literature review on the effect of IHDs on

Ito et al. (2018) and Gillan (2017) demonstrate that moral suasion significantly induces energy conservation. In addition, the evidence that social comparison information used in this study promotes environmental pro-social behaviors has been provided by several social psychological studies (Festinger, 1954; Schultz, 1999; Kurz et al., 2005; Schultz et al., 2007; Goldstein et al., 2008; Nolan et al., 2008).[4]

Only two recent studies have directly compared the treatment effects of monetary and nonmonetary incentives in a unified field experiment. Ito et al. (2018) found that nonmonetary incentives reduced peak-hour electricity usage by 8% for the first three event days, but the effect reduced to zero in a repeated intervention. However, a 14%-17% reduction in peak-hour electricity usage from monetary incentives (CPP treatments) was observed during the whole experimental period. Similarly, Gillan (2017) found that the consumption reduction induced by nonmonetary incentives was not significantly different from zero, while the reduction from monetary incentives (CPP treatments) was equal to 6%. The nonmonetary incentive utilized in both studies was moral suasion (instead of social comparison). To the best of our knowledge, a direct comparison between monetary incentives and social comparison nudge using a unified field experiment has not been carried out yet in the economics literature. Further, the assessment and comparison of the heterogeneity of the treatment effects among households have not been conducted. Thus, this study is the first to directly compare and investigate the differences between the treatment effects of nonmonetary incentives (social comparison nudge) and conventional monetary incentives by using a unified field experiment.

Leading studies on social comparison and energy conservation behaviors have involved field experiments using the OPOWER's Home Energy Report (HER) (Allcott and Mullainathan, 2010; Allcott, 2011; Ayres et al., 2012; Costa and Kahn, 2013; Allcott and Rogers, 2014; Allcott and Kessler, 2019; Knittel and Stolper, 2019). Overall, they show that the average treatment effect (ATE) of sending bimonthly HER with social comparison information is a 2% monthly reduction in electricity consumption. At the same time, the effects are heterogeneous: households in the highest decile of pre-treatment consumption reduce usage by 6.3%, while consumption by the lowest decile of customers reduces by 0.3% (Allcott, 2011).[5] Furthermore, Allcott and Kessler (2019)

---

electricity consumption.

[4] Social comparisons have been investigated in many field experiments and across a broad spectrum of behaviors, including recycling behavior (Schultz et al., 1999), water and energy conservation (Kurz et al., 2005), electricity conservation (Schultz et al., 2007; Nolan et al., 2008), and indirect water conservation (i.e., reuse of towel in hotels) (Goldstein et al., 2008).

[5] For another example of heterogeneity, Costa and Kahn (2013) show that liberal

show that the estimated HER welfare gains might be overstated by the “moral utility” of the recipients.[6]

To explore a nonmonetary incentive’s effectiveness as a public policy tool, its potential heterogeneity needs to be addressed. Recent developments in machine-learning techniques allow us to estimate the heterogeneous treatment effects for each household. The large number of observable characteristics provides this technique with more attractive applications (Athey and Imbens, 2017). Although conventional machine learning has been used for predicting outcomes using observable variables instead of estimating the parameters of the treatment effects, more recent studies have often utilized these algorithms for causal inference. In the context of energy conservation, Knittel and Stolper (2019) explore the heterogeneous treatment effect of social comparison information in OPOWER’s HER using causal forest (CF) (Wager and Athey, 2018) expanded from a causal tree (Athey and Imbens, 2016), thus estimating conditional average treatment effect using a regression tree.[7] Knowing different responses from individuals in different subgroups may increase policy efficiency. For example, Allcott and Kessler (2019) suggest that substantial welfare improvement is expected by applying selective targeting based on the heterogeneous willingness to pay for HER with social comparison. We demonstrate that the selective targeting according to the heterogeneity estimated by the CF algorithm improves treatment effects.

This study makes three primary contributions to the literature. First, our experimental setup is the first to directly compare monetary incentives (rebate) and nonmonetary incentives (social comparison nudge) in a cohesive field experiment, shedding light on the heterogeneity of treatment effects. We focus on electricity conservation in peak-demand hours, in which the marginal cost of electricity is substantially higher than the rest of the day. Second, we analyze the heterogeneous treatment effect for each household to reveal the mechanisms behind the heterogeneity by using the machine learning technique. Finally, we suggest a novel feasible targeting strategy based on a rich set of household characteristics. As part of the study, we conducted an RCT on 954 households with advanced electricity meters, often called smart meters, in Japan. We randomly assigned all households to one of the three groups: Control (C), Rebate (R), and Nudge

households reduce energy usage in response to HERs two to four times more than conservative ones.

[6] Glaeser (2006) fears a “psychological tax” or “moral tax” of a nudge, which reduces consumer welfare, while it does not generate revenues.

[7] The first study utilizing causal forest algorithms to evaluate treatment effects in RCT is Davis and Heller (2017b), which examined youth employment problems in the summer.

(N), and observed their behavioral changes in terms of electricity usage. Using high-frequency data (30-minute intervals) of household electricity usage and observable household characteristics, we estimated the average treatment effects (ATEs) and heterogeneous treatment effects at the household level and the determinants of such heterogeneities. Moreover, we assessed the improvement in the expected treatment using optimum targeting strategies based on the heterogeneous treatment effects predicted by a machine learning technique.

Our main findings are as follows. Using a difference-in-difference (DID) regression approach, we first estimate the ATE. We find that the rebate intervention reduces electricity consumption by 4%, and the social comparison nudge intervention lowers electricity consumption by 1%, but such an effect is not significantly different from zero. Second, by applying a novel machine learning approach for causal inference (CF) to estimate heterogeneous treatment effects at the household level, we demonstrate that the treatment effects of the nudge intervention are characterized by twice as large heterogeneity among households than in the case of the rebate intervention. Moreover, while the effects of the rebate intervention (i.e., boosting conservation) are intended for all households, in the case of the nudge intervention, 37% of households show an unintended reaction (i.e., increasing consumption). The proposed CF methods also reveal that these heterogeneities depend on a household's electricity use and their differences with respect to similar households. Third, we demonstrate that selective targeting, where the nudge intervention is allocated to customers for which the predicted treatment effect of a nudge is larger than that of a rebate and the rebate intervention is allocated to those for which the predicted treatment effect of a rebate is larger than that of a nudge, is expected to improve the ATE, generating a 6% reduction in electricity use.

The remainder of this paper is organized as follows. Section 2 explains the experimental design and treatment and describes the data, and Section 3 introduces the DID regression model and results on the ATEs. The CF method and results are presented and discussed in Section 4. Section 5 discusses findings and the implications of the study. Section 6 provides the conclusions.

2. Experimental Design, Treatment, and Data

2.1. Experimental Design and Data

From November 2019 to February 2020, we conducted a field experiment targeting the customers of Chubu Electric Power Co., Inc., which has the second-largest share of the retail electricity market in Japan. We contacted more than 20,000 residential electricity customers in the Chubu Electric Power area through a Japanese internet research company.[8] All customers were informed about the project aim of saving residential electricity consumption using information from advanced electricity meters, survey schedules, and participation rewards. We discarded households that were not customers of Chubu Electric Power, had planned to move residence during the experiment, or did not have smart meters. Overall, 1,168 customers confirmed their participation. We excluded customers whose electricity use data over the experiment period were not technically obtainable and those for whom we could not determine whether they received interventions. The final sample comprises 954 households.

We randomly assigned the 954 households to one of three groups: control (C), rebate (R), and nudge (N).[9]

Control Group (C): The 327 households in this group received information feedback about daily electricity use (kWh) and peak-time electricity use (kWh) and a participation reward. They received no other intervention.

Rebate Group (R): The 313 households in this group received information feedback about daily electricity use (kWh) and peak-time electricity use (kWh) and a participation reward. In addition, they received monetary incentives for energy conservation, as described below.

Nudge Group (N): The 314 households in this group received information feedback about daily electricity use (kWh) and peak-time electricity use (kWh) and a participation reward. In addition, they received a social comparison nudge for energy conservation, as described below.

The experiment was implemented in three phases, summarized in Figure 1. In the first phase (the first survey), we collected household-level electricity usage data at 30-minute intervals in the early winter (December 1–14)[10] as well as household characteristics (e.g.,

---

[8] MyVoice (https://www.myvoice.co.jp/)

[9] Before confirming the experimental design, we conducted a pre-experiment for a random sample of households different from the recipients of our main survey. In the pre-experiment, we randomly assigned households to one of four groups: control (C), rebate (R), nudge (N), and rebate + nudge (RN), and examined the combined effect. Based on the estimated ATE and standard error in the pre-experiment, we excluded the fourth group (RN) due to the lack of crowd-out effect and synergy effect in combination.

[10] After electricity deregulation, from April 2016, electricity customers with a smart-meter have had the ability to autonomously check and download their high-frequency

monthly electricity bill, electricity rate plan, consciousness to save power, and number of appliances and units).[11] In the second phase, households were stratified and randomly assigned to one of the above three groups based on daily and peak-time consumption data during the early winter and received different energy reports (January 24–30). The treatment groups also received a notification about rewards or a nudge for energy conservation during seven peak-load event days (January 31 – February 6). Households in the treatment groups received a pre-event e-mail on the day before the peak-load event (at 5 pm, January 30). In the third phase (the second survey), after the seven event days, we collected household-level electricity usage data at 30-minute intervals from January 1 to February 6, including pre-period data and event period data,[12] and asked customers their impression regarding the energy reports. At the end of the experiment, we provided feedback and a participation reward of 1,000 yen (approximately $9 in 2020): 300 yen for responding to the questionnaires and 700 yen for completing the entire survey process.

＜Figure 1. Timeline and procedures for the experiment＞

Table 1 presents the summary statistics of pre-experiment consumption data and demographic variables by group. Due to the stratified randomization based on electricity use (i.e., daily use and peak-time use) in the "Early Winter" season, the observables are balanced across groups.[13] Electricity usage during the early winter, pre-event period, and rebate baseline period as well as other electricity-related demographic characteristics are similar across groups.

We collect household-level electricity consumption data of 30-minute intervals in peak-demand hours (5 pm – 9 pm) from January 1 to February 6, discarding the seven days between January 24 and January 30 (before the event days) because the households received group-specific energy reports during this period. We only assessed the behavior of households that received the energy report for certain by ascertaining their response to

data. Using this system, we asked participants to download 30-minute interval data from their member pages and send them without their private information (i.e., name, residential address). In the first survey, we obtained 14 days of data from December 1 to December 14.

[11] The questionnaire is available upon request.

[12] In the second survey, we obtained 37 days of data from January 1 to February 6.

[13] We used stratified randomization to assign households to each group based on the daily electricity consumption (three blocks) and peak-time (four blocks) because different consumption patterns were observed in daily use and peak-time in the pre-experiment.

simple questions submitted via a quick survey.[14] Thus, using high-frequency data on household electricity usage from January 1 to January 23 as pre-event data and the data from January 31 to February 6 as event data, we examine how the proposed treatments affect electricity usage in peak-demand hours for the 954 selected households.

Winter weather patterns are typically observed from late November until early March in the experimental region. Because customers experience the coldest time of the year from December to February in this region, winter patterns (when residents use heating appliances) in electricity usage are observed during our experimental term.[15] We randomly assigned households to groups based on "Early Winter" electricity consumption and used the consumption data of January and February to examine the treatment effects of interest.

＜Table 1. Summary statistics by control group and treatment groups＞

2.2. Treatments

A. **Monetary Incentive**

Our first treatment is a monetary incentive, a rebate for energy conservation. We provided information feedback about daily electricity use (kWh) and peak-time electricity use (kWh) for the control group and the treatment groups (Figure 2). Y_i^{base} indicates the rebate baseline consumption of household i, which is the total electricity use in peak-demand hours of the week (January 17–23) before informing the households of the rebate rule. Y_i^{event} indicates the consumption of household i, which is the total electricity use in peak-demand hours of the event week (January 31– February 6). We defined total electricity conservation in peak-demand hours per week, ΔY_i^{total}, as the difference between Y_i^{event} and Y_i^{base} for household i. Thus, $\Delta Y_i^{total} = -\min\{0, Y_i^{event} - Y_i^{base}\}$ because household i obtains zero incentive if energy conservation is negative

[14] For the excluded households, we tested for balance in the observable characteristics across groups and confirmed statistical balance in observables. Thus, no different features were observed across groups for the excluded households.

[15] We compared the monthly average high and low temperatures between the Nagoya prefecture in Japan (the mid-part of our experimental region) and Washington DC in the US. We provide this comparison in Figure A1 in the Appendix. Little variety is observed in weather patterns in Japan, except for the Hokkaido prefecture. The average low and high temperatures are very similar between the three largest retail electricity market areas (Kanto, Chubu–our experimental region, and Kinki) in Japan and Washington DC.

$(Y_i^{event} - Y_i^{base} < 0)$.

Therefore, the total amount of rebate per event week for household i can be calculated as follows:

$$Q_i = \min\{\Delta Y_i^{total} * 100 \text{ JPY}, 1000 \text{ JPY}\}. \qquad (1)$$

The households in the rebate group (R) received 100 yen per 1 kWh conservation if electricity use in peak-demand hours of the event week (January 31–February 6) was less than that of the previous week (January 17–23). The maximum rebate was 1,000 yen.[16] The households received rebates in addition to their participation rewards.

To secure the transparency of the incentive design, we informed households in the rebate group (R) of the rebate calculation rule before the event days. The messages sent to this group after the group assignment were "Substantial energy conservation will be required for the society during the coldest part of winter (January and February in particular)" and "Please reduce your electricity usage in the critical peak-demand hours during the following week." In addition, messages sent also included "During 5 pm – 9 pm (peak-demand hours) of the event week from January 31 to February 6, you will receive rebates for your electricity conservation in addition to your participation rewards" and "Rebate will be 100 yen per kWh (maximum 1,000 yen in total per week)," with notes saying "Your total electricity conservation will be calculated as the reduction from your usage of the normal week of January 17–23."[17]

The households received these messages between January 24 and January 30. Thus, they were unable to manipulate their rebate baseline consumption of the normal week (January 17–23). While baseline-based rebates provide customers with incentives to reduce electricity use during event days, they may also create undesired incentives for customers to manipulate their baseline consumption (Wang and Tang, 2018). This "baseline manipulation" has often been observed in RCT field experiments (Wolak, 2007)

---

[16] We confirmed this maximum rebate prompted by budget constraints. The maximum rebate amount of 1,000 yen corresponds to 10 kWh electricity conservation in peak-demand hours. The average electricity use in peak-demand hours of the pre-event period was 0.43 kWh/30-min., and the total use in seven days was approximately 24 kWh. Thus, the maximum rebate amount (1,000 yen; 10 kWh equivalent) corresponds to more than 40% (drastic) reduction from normal use. In our experiment, the average rebate amount was 145 yen for all households in the rebate group (R), and seven households (2% of these) received 1,000 yen. In addition, the electricity rate during these hours is 21–28 yen per 1 kWh in normal conditions. These elements confirm that the maximum rebate (1,000 yen) and rebate unit (100 yen per 1 kWh) are sufficient incentives for conservation activities in this experiment.

[17] These messages are all provided in Figure A2 in the Appendix.

and causes significant errors in the prediction of demand and treatment effects. However, in the proposed experiment, strategic manipulation was impossible because the rebate calculation rule was not announced until January 24. As shown in Table 1, the rebate baseline consumption (January 17–23) is statistically similar across groups.

We provided a list of familiar energy-saving activities and the consequent electricity consumption reduction (kWh) for the rebate group.[18] To ensure that participants could easily access the information about these energy-saving activities, event days and hours, and rebate rules, we sent a one-page summarized document (via e-mail). At the end of the document, we added a QR code for detailed information on the energy-saving activities provided by the Japanese Agency for Natural Resources and Energy so that participants could learn more about energy-saving actions. We delivered notifications to customers at 5 pm on January 30 by sending a text message to their cell phones or an e-mail to their computers with the URL of the one-page summarized document mentioned above.

B. **Nonmonetary Incentive**

Our next treatment is a nonmonetary incentive, which is a nudge based on social comparison information. After assigning customers to the nudge group (N), we provided them with information feedback about daily electricity use (kWh) and peak-time electricity use (kWh) in line with the other groups and informed them of the social comparison reports, as follows. The key feature of the social comparison information is a bar graph comparing a household's electricity use during peak hours in the early winter period to the mean electricity use by similar households (Figure 3). The latter is the predicted value for each household obtained through the Random Forest (RF) approach (Breiman, 2001) based on 81 observable characteristics identified in the first survey, which addressed electricity use in 30-minute intervals, monthly energy bills, demographic characteristics, lifestyle characteristics related to energy consumption, and owned appliances and number of units.[19] The households that use less electricity compared to

---

[18] The list of activities consists of six familiar energy-saving actions: efficient use of air conditioner, refrigerator, TV, lights, and stand-by power, and enjoying outside activities. We also indicate how much electricity (kWh) is possibly saved in seven days by committing to each action.

[19] The algorithm that identifies "similar" households in the O-power HER is not disclosed, but it is a function of 100 geographically nearest neighbors in similar house sizes (Allcott and Kessler, 2019). In this study, we predicted electricity use during peak hours for each household by using RF based on the 81 characteristics obtained in the first survey as "similar" households' electricity usage.

similar households earn a “good!” mark, which represents an “injunctive norm” and is expected to reduce the boomerang effect. However, we did not provide any mark related to the injunctive norm (i.e., bad mark) to households with high relative use to avoid customers’ dissatisfaction.[20]

To secure the transparency of incentive design, we informed households in the nudge group (N) that they would receive feedback about their energy conservation achievements at the end of the season. The messages sent to this group after the group assignment were “Substantial energy conservation will be required for the society during the coldest part of winter (January and February in particular)” and “Please reduce your electricity usage in the critical peak-demand hours (5 pm – 9 pm) of the event week from January 31 to February 6.” There were also messages such as “You will receive feedback information on how much electricity you have conserved,” with notes saying “Your total electricity conservation will be calculated as the reduction from your usage of the normal week of January 17–23.”

We provided a list of familiar energy-saving activities and the consequent amount of electricity reduction (kWh) and CO2 reduction (kg-CO2) to the nudge group. To ensure that participants could easily access the information about these energy-saving activities and event days and hours, we also sent them an e-mail with a one-page summarized document. At the end of the document, we added a QR code for detailed information on energy-saving activities, similarly to the rebate group so that participants could learn more about energy-saving actions, if interested. We delivered notifications to customers at 5 pm on January 30 by sending a text message to their cell phones or an e-mail with the URL of the above one-page summarized document.[21]

＜Figure 2. Energy report for the control and rebate groups＞

---

[20] The earlier versions of O-power HER included an injunctive norm (smiley faces), which could eliminate the boomerang effects (Allcott, 2011). More recent versions of the HER did not use any injunctive norm; hence, boomerang effects were observed (Knittel and Stolper, 2019).

[21] OPOWER’s HER comprises two features: a bar graph comparing a household’s energy use to that of its neighbors and energy conservation tips suggesting ways to achieve energy saving by daily behavioral changes and replacing durable appliances. Our energy report follows these features (i.e., social comparison and conservation tips); we also provide information regarding how to read the hourly consumption data graph visualized in the customers’ member pages (e.g., the correlation between the usage and the hourly behaviors). Hence, customers can learn more about the relationships between their regular specific activities and electricity consumption related to their interests.

＜Figure 3. Energy report for the nudge group＞

3. DID Estimation

3.1.Estimation of ATEs

As mentioned above, we use household-level electricity consumption data of 30-minute intervals in peak-demand hours of the pre-event and event periods to estimate the ATE, as follows:

$$\ln Y_{it} = \sum_{d \in \{nudge, rebate\}} \beta_d Z_{it}^d + \theta_i + \lambda_t + \eta_{it}, \qquad (2)$$

where Y_{it} is the electricity usage of household i in 30-minute interval t. Z_{it}^d equals one if household i is in group R (rebate) or group N (nudge) and receives treatment in t. We included household fixed effect θ_i, and time fixed effects λ_t for each 30-minute interval to control for time-specific shocks (such as changing weather conditions). We clustered the standard errors at the household level to adjust for serial correlation. η_{it} is an unobserved error term with mean equal to zero. β_d captures the ATE of each treatment. We used the natural logarithm of electricity usage for the dependent variable, $\ln Y_{it,}$ so that we may approximately interpret the treatment effects in percentage terms.

3.2. DID Estimation Results: ATE

Table 2 reports the ATEs estimated using the proposed DID regression model. Column 1 of Table 2 shows that the rebate intervention (100 yen per 1 kWh) caused a significant reduction in peak-hour electricity usage (4.3% for the treatment days), and the social comparison nudge caused a 0.7% reduction, but such an effect is insignificant. Columns 2 and 3 of Table 2 show that the ATEs estimated using split samples are highly heterogeneous as significant differences exist between lower users (Less) and higher users (More) during peak-demand hours. Note that lower users (higher users) are defined as households whose peak-time electricity consumption is lower (higher) than the usage predicted by the RF using observable characteristics and consumption data in 30-minute intervals, both obtained in the first survey.[22] For the households in the nudge group, we

[22] Of the 954 households, 18 received the note “You use approximately the same amount of electricity as similar households,” as their consumption did not differ from the RF-predicted usage by more than 1 percent. Thus, the split samples used for estimation in Columns 2 and 3 do not include these 18 households.

provided messages on lower (higher) electricity use than similar households for lower users (higher users) as social comparison information. Thus, these ATEs capture the behavioral changes motivated by each social comparison nudge. For the households in the rebate group to whom we did not provide the above information, the ATEs demonstrate heterogeneity in the behavioral changes motivated by the rebate intervention based on the patterns of peak-time electricity consumption.[23] Column 2 shows that the ATEs for lower users are significant and larger in both interventions: −5.6 % for the rebate and −3.8% for the nudge. However, the results in Column 3 indicate that the ATEs for higher users are not significantly different from zero in either intervention.

＜Table 2. ATE of rebate and nudge＞

According to previous studies, monetary incentives induce 7%–22% electricity reduction during peak-demand hours in pricing-style interventions (CPP) (Jessoe and Rapson, 2014; Ito et al., 2018) and less than a half of that in rebate-style interventions (CPR or critical peak rebate) (Wolak, 2011). These findings are consistent with the result of our study (4.3%). Regarding nonmonetary incentives, only two studies investigate the treatment effects of a nudge for energy conservation during peak-load events (Ito et al., 2018; Brandon et al., 2019). Ito et al. (2018) utilize moral suasion, and Brandon et al. (2019) utilize peak-time energy reports (PER) with social comparison as a nudge intervention. The latter is the closest analog to our study. They address three peak-load events during two months in the summer (August and September) of 2014 and estimate the ATEs of PER with social consumption information on peak-time electricity use using hourly consumption data.[24] They find that the ATE of PER for households that have not received bimonthly HER is −3.8%, in line with our estimate for households with low relative use. Unlike their experiment, however, we address a continuous seven-day event, which causes a relatively smaller treatment effect, on average, as customers are expected to make continuous efforts or intermittent efforts choosing several easy days to conserve electricity use during these seven days. Ito et al. (2018) address intermittent 21-day peak-load events during the winter of 2013 and show that the effects of their nudge (i.e., moral

[23] Table B2 in the Appendix compares the covariates across subgroups. For each subgroup, we tested for balance in observable characteristics between the control group and the nudge group, and we confirmed statistical balance in observables. Thus, no different features were observed across subgroups.

[24] Their interventions were conducted via telephone call or a combination of telephone call and e-mail notification on the afternoon of the day immediately before a peak event. Event days occurred one time in August and two times in September.

suasion) quickly shift toward zero, while their nudge causes an 8.3% reduction in peak-time electricity use for the first three days. Their findings indicate that customers find it hard to consistently commit to electricity conservation.

Although heterogeneity in treatment effects based on the relative daily electricity consumption of households has been observed for bimonthly HER (Allcott, 2011), no heterogeneity has been detected for PER based on baseline electricity consumption (Brandon et al., 2019). In this study, we observe treatment effect heterogeneity based on peak-time electricity consumption: the ATEs of rebate and nudge are −5.6% and −3.8%, respectively, for households with lower peak-time consumption than the peak-time usage predicted by RF and are not significantly different from zero for households with higher peak-time consumption.

A social comparison message provides "upward social comparison" to households with higher consumption and "downward social consumption" to households with lower consumption. It should be noted that an upward social comparison may encourage efforts to conserve energy in some cases but may also discourage consumers when they feel inferiority or frustration in strongly self-related activities. In this study, on the one hand, the latter negative influence seems significant in the short term. On the other hand, downward social comparison is found to stimulate positive feelings, such as superiority or happiness, promoting positive impacts and increasing people's performances, according to a large body of social and psychological studies. Such positive feelings were strongly stimulated by the injunctive norm (i.e., the "good!" mark) in our experiment.[25]

4. Causal Forest

4.1. Estimation Model

We estimate heterogeneous treatment effects (HTE) by using the CF approach (Wager

---

[25] Our follow-up survey data (i.e., response data in the second survey) indicate that social comparison messages induce positive feelings (feeling inspired or proud) for lower users and negative feelings (feeling pressured or guilty) for higher users. Regarding households with higher electricity use, 26% stated that the HER made them feel "pressured," and 4% declared that it made them feel "guilty"; both of these values are 1.5 to 2 times greater than those of the households with lower electricity use. Regarding the latter, 59% stated that the HER made them feel "inspired," and 10% declared that it made them feel "proud"; both values are 1.1 to 10 times greater than those of the households with higher use. For a comprehensive review of the evidence related to the effects of social comparison on human behaviors in social psychology, see Bunk and Gibbons (2007).

and Athey, 2018). The CF allows us to estimate treatment effects conditional on household characteristics and, therefore, to predict household-specific treatment effects. With the growing interest in the HTE, economists have begun applying the CF in various contexts such as youth employment programs (Davis and Heller, 2017a, 2017b), interventions to induce energy saving (O'Neill and Weeks, 2018; Knittel and Stolper, 2019), and rural development programs (Carter, Tjernström, and Toledo, 2019).

The CF is built on the RF, which solves regression and classification problems by aggregating *B* trees grown with *B* bootstrap subsamples into a forest. Each tree grows by splitting a variable at a certain value in each node. More precisely, the algorithm chooses *m* household characteristics from the set of *p* ($m < p$) characteristics and splits one of the *m* characteristics at each node so that the variance of an outcome is maximized after each split (Hastie et al., 2009; Biau and Scornet, 2016). In this study, we apply the Generalized Random Forest (GRF) algorithm (Athey, Tibshirani, and Wager, 2019), which comprises the CF algorithm as a special case.

Let $\tau^d(x)$ denote the HTE, the parameter of interest in this study, which can be defined for each treatment $d \in \{\text{nudge, rebate}\}$ as follows:

$$\tau^d(x) = E\left[Y_i\left(Z_i^d = 1\right) - Y_i\left(Z_i^d = Z_i^{d'} = 0\right) \,\middle|\, X_i = x \right], \qquad (3)$$

where $d' \neq d$, and $Y_i\left(Z_i^d = 1\right)$ and $Y_i\left(Z_i^d = Z_i^{d'} = 0\right)$ denote potential outcomes realized if household *i* receives treatment *d* and if household *i* does not receive any treatments, respectively. In other words, $\tau^d(x)$ represents the treatment effects conditional on household characteristics. For households with $Z_i^d = 1$ or $Z_i^d = Z_i^{d'} = 0$, the GRF calculates $\tau^d(x)$ as the solution of the following moment equation:

$$E\left[\psi_{\tau^d(x), c^d(x)}\left(Y_i, Z_i^d\right) \mid X_i = x \right] = 0, \qquad (4)$$

where $c^d(x)$ is a constant term, and $\psi(\cdot)$ represents a score function. Under the GRF framework, the score function is specified as follows:

$$\psi_{\tau^d(x), c^d(x)}\left(Y_i, Z_i^d\right) = \left(Y_i - \tau^d(x) \cdot Z_i^d - c^d(x)\right)\begin{pmatrix} 1 \\ Z_i^d \end{pmatrix}. \qquad (5)$$

The GRF algorithm grows *B* trees, as the RF algorithm does. Leaves, the edges of each tree, contain information on leaf-level treatment effects. The HTE is calculated as the weighed sum of such leaf-level treatment effects. Trees are grown in a way that maximizes the heterogeneity of treatment effects after each split.[26] The GRF algorithm

[26] The GRF algorithm uses a criterion that approximates the heterogeneity to reduce the

proceeds as follows:

1. Take S random subsamples from the whole sample without replacement;
2. Divide the subsamples into two sets (S_1 and S_2);
3. Use S_1 to grow trees;
4. Use S_2 to compute $\tau^d(x)$;
5. Iterate steps 1–4 for B times;
6. Weight trees to aggregate into a forest.[27]

The HTE computed above is asymptotically consistent and normal. In this study, we use the R package *grf* (Tibshirani et al., 2018) to estimate the HTE.

4.2. Empirical Specification

To implement the GRF algorithm, we use 30-minute interval electricity usage in peak-demand hours during the pre-event and event periods divided by the mean households' electricity usage during the pre-event period as an outcome. While this outcome has a panel structure, the package *grf* does not provide an algorithm for panel data. Hence, we try to capture the features of panel data as follows. First, to control for time trends, we include in our covariate set discrete variables that increase at each 30-minute interval or each day. Note that these variables are not dummies but variables that grow over time. Since we focus on the peak-demand hours, the 30-minute interval variable takes one of eight values (1,…, 8). The day variable takes one of 30 values (1,…, 30) as we use data for the pre-event (from January 1 to January 30) and event periods (from January 31 to February 6), excluding the energy report period (from January 24 to January 30). Second, we use the "cluster" command to cluster households. As a result, observations from the same household are more likely to be selected in the first step by the GRF algorithm (Athey and Wager, 2019). By clustering, we can incorporate correlations within households into the proposed estimation model.

Additionally, we use 86 household characteristics obtained from the smart meter data and survey. We construct five variables using these smart meter data. The first is the difference in energy consumption in the peak-demand hours compared to that of "similar households" in early winter (see Section 2.2). The other four are the mean, standard

computational burden.

[27] The weights are computed inside the algorithm. The weight for household i is determined based on how often household i belongs to the same leaf as x.

deviation, maximum, and minimum of energy consumption during the pre-event period. From the survey data, we construct variables for the socio-economic characteristics of customers such as their age or income, building-related characteristics such as the age of the house or home size, and electricity-usage-related characteristics such as the number of durable electric goods or the electricity plan. Note that this study is primarily interested in finding variables that are useful for targeting purposes. Hence, we follow Athey and Wagner (2019) to automatically select essential variables within the algorithm.

Table 3 reports the ATEs during the event period as calculated by the GRF algorithm. The results are consistent with those presented in Column 1 of Table 2, even though the absolute values are slightly larger. On average, rebate intervention reduces energy consumption more than nudge intervention. The former exhibits statistically significant effects, while the effects of the latter are indistinguishable from zero. These results suggest the validity of the proposed specification for the GRF algorithm.

<Table 3 ATEs computed by the GRF algorithm>

4.3. Heterogeneous Treatment Effects

Figure 4 plots the distributions of household-specific treatment effects predicted by the GRF. The horizontal axis measures the treatment effects, while the vertical axis represents the fraction of each bin. Panels A and B represent the treatment effects of the rebate and nudge interventions, respectively. The treatment effects measure the reduction rate of 30-minute consumption relative to the mean 30-minute consumption during the pre-event period.

<Figure 4 Distributions of heterogenous treatment effects>

Figure 4 indicates that both interventions generate heterogeneity in treatment effects. However, the degree of heterogeneity differs across interventions. The predicted treatment effects of the rebate intervention are negative (i.e., reducing electricity usage) for all of the households. The standard deviation of the distribution is 1.5%, and the minimum and maximum are −8.5% and −0.79%, respectively. However, the treatment effects of the nudge intervention are more heterogeneous than those of the rebate intervention and positive (i.e., increasing electricity usage) for 36.5% of the households. The standard deviation is 3.8%, with a minimum of −17.9% and a maximum of 15.0%. These results suggest that the treatment effects may be improved by predicting the treatment effect heterogeneity, especially that of the nudge intervention, and by assigning households to an intervention that leads to larger treatment effects. If observable household characteristics can largely explain heterogeneity, targeting based on such characteristics may improve the efficiency of a policy.

Figure 5 represents the "Variable Importance," which measures the fraction of times a household characteristic is used for splits. We can interpret variables with high Variable Importance as essential determinants of the treatment effect heterogeneity.

<Figure 5 Key variables for growing trees>

Figure 5 shows that electricity-usage-related characteristics are fundamental in both interventions. The difference in electricity usage compared to that of similar households in early winter ranks the highest. In addition, the mean, standard deviation, maximum, and minimum of consumption during the pre-event period also rank high. The treatment effects are heterogeneous over electricity-usage-related characteristics, in line with the results of O'Neill and Weeks (2018), who investigate the effects of economic incentives on energy saving, and Knittel and Stolper (2019), who investigate the effects of non-economic incentives. These results imply that if a policymaker can observe electricity-usage-related characteristics, especially the difference in consumption relative to similar households, the treatment effects may be improved via targeting. We further discuss this point in the next section.

In addition to electricity-usage-related characteristics, Figure 5 also confirms other findings in the literature. For example, O'Neill and Weeks (2018) show that the number of laptops is an essential determinant of heterogeneity. Moreover, the existence of an association between treatment effect heterogeneity and household characteristics, such as solar panel possession, income, household size, and age, is in line with Ida, Murakami, and Tanaka (2016), who investigate the treatment effect heterogeneity of dynamic pricing on energy saving. While their study does not find a significant association between the number of air conditioners and home size, Figure 5 shows that these characteristics are essential for growing trees. Finally, the mean electricity and gas bills and house age, which have not been investigated in the literature, also seem to play a vital role.

4.4. Slope Test of Heterogeneity

To examine whether the predicted treatment effects estimated in Section 4.3 reflect true treatment effect heterogeneity, we adopt the sample splitting approach proposed by

Chernozhukov et al. (2019), also applied by Davis and Heller (2017b) and O'Neill and Weeks (2018). We first divide the whole sample into training and test samples. Then, we estimate a forest using the training sample and predict treatment effects $\hat{\tau}^d$ with the test sample. Based on these predictions, we run the following regression and save its coefficients and standard errors:

$$Y = X\beta^d + \gamma^d(Z^d - p^d) + \eta^d(Z^d - p^d)\left(\hat{\tau}^d - E(\hat{\tau}^d)\right) + \varepsilon, \qquad (6)$$

where p^d is the propensity score. We weight the regression by $1/(p^d(1 - p^d))$. After iterating this procedure 1000 times, we calculate the median of the estimates and confidence intervals of size $\alpha/2$. Chernozhukov et al. (2019) show that this algorithm provides point estimates and their $(1 - \alpha)$ confidence intervals. Specifically, if η^d is statistically distinguishable from zero, we can conclude that there exists heterogeneity in treatment effects, and our predictions are relevant to the true heterogeneous treatment effects. In addition, γ^d measures the expected treatment effects. Table 4 summarizes the estimation results of the sample splitting approach.[28]

<Table 4 Results of the slope test>

Table 4 shows that the results of ATE (γ^d) are consistent with those reported in Column 1 of Table 2. Namely, the ATE of the rebate intervention is −4.4% and is statistically significant, while that of the nudge intervention is −0.8% and is statistically indistinguishable from zero. Importantly, the results in Column 2 of Table 4 show that in both interventions, η^d is statistically different from zero. Overall, the results of the sample splitting approach indicate that the treatment effects of both rebate and nudge interventions are heterogeneous, and our predicted values based on the GRF approach capture the true treatment effect heterogeneity.

5. Improvement of Treatment Effects Through Targeting

The existence of treatment effect heterogeneity in both interventions and the difference in their distributions suggest that a preferable intervention in terms of its impacts on energy saving differs from household to household. Therefore, policy designs that selectively assign one of the interventions to each household would improve treatment

[28] For implementing the sample splitting approach, we modify the code by Davis and Heller (2017b).

effects and welfare gains (Allcott and Kessler, 2019). In this section, we leverage the estimated treatment effect heterogeneities and a rich set of household characteristics to develop a novel feasible targeting strategy. To this end, we compare predicted treatment effects of the monetary intervention to those of the nonmonetary intervention at the household level and assign an intervention that generates larger treatment effects.

Figures 6 and 7 intuitively show how targeting improves treatment effects. The horizontal and vertical axes measure the values of household characteristics and the predicted treatment effects, respectively. The solid lines represent local polynomial regression curves. The household characteristics in Figures 6 and 7 are the mean electricity usage during the pre-event period and the difference in energy consumption compared to similar households in early winter, respectively. Panels A and B report the results of the rebate and nudge interventions, respectively.

<Figure 6 Mean electricity usage in the pre-event period and treatment effects>

<Figure 7 Difference in electricity usage and treatment effects>

Both panels in Figures 6 and 7 confirm the existence of heterogeneity in the treatment effects. Panel A of both figures shows that the treatment effects of the rebate intervention decrease in both mean energy consumption and the difference in consumption relative to similar households. The treatment effects of the nudge intervention exhibit a similar pattern, as shown in Panel B. However, unlike the rebate intervention, the local polynomial regression curves of the nudge intervention turn positive at certain thresholds: 0.5kWh for the mean consumption during the pre-event period and 0kWh for the difference in consumption compared to similar households in early winter. To improve the treatment effects, households whose characteristics are above these thresholds should be assigned to the rebate intervention rather than the nudge intervention. In contrast, households whose values lie in the leftmost region (i.e., low mean usage and difference) have larger treatment effects and should be assigned to the nudge intervention. Thus, targeting based on these characteristics can improve the treatment effects.

The negative correlation between mean energy consumption and treatment effects contradicts O’Neill and Weeks (2018) and Knittel and Stolper (2019). They find that the treatment effects increase in mean energy consumption. This difference in results may be attributed to the diversity in the research designs. This study focuses on the peak-time energy consumption observed in a short intervention period (i.e., about one week), while their findings are based on daily energy consumption observed over a longer period. The long period intervention allows households to reduce energy consumption by (i) replacing durable electric goods, (ii) changing their behavior regarding how to use existing goods, or (iii) both. In contrast, the short period intervention is likely to only encourage the second effect. Compared to behavioral changes, replacing durable appliances can result in a larger reduction in energy consumption. The fact that households with high mean electricity usage are more likely to replace appliances in the case of a long-period intervention can explain their significantly different results.

With the estimated forests, we can predict the household-specific treatment effects of both interventions, which determine which intervention is preferable for each household in terms of treatment effects. Columns 1 and 2 of Table 5 represent the treatment effects realized if all households are uniformly assigned to one intervention and intervention is assigned only to households with negative treatment effects, respectively. The first and second rows of Columns 1 and 2 show the results of the rebate and nudge interventions, respectively. In addition, the third row presents the treatment effects realized if an intervention with larger effects is assigned. Note that the forest of the nudge intervention is separately estimated after splitting the nudge and control groups into three different subsamples. As discussed in Section 3, the impact of information regarding the difference in energy consumption relative to similar households on energy saving depends on whether a household consumes more or less electricity relative to similar households (see Table 2). To capture this difference, we divide the sample into three subsamples according to whether a household consumes more, less, or the same as similar households, and we estimate three corresponding forests using the GRF algorithm.

<Table 5 Improvement in treatment effects by targeting>

“Uniform” and “Targeting” of the rebate intervention coincide at −5.02％ because, as shown in Panel A of Figure 4, the rebate intervention results in negative effects for all households. In contrast, as the nudge intervention generates larger heterogeneity, by targeting only those households whose predicted treatment effects are negative, we can improve the treatment effects from −1.25％ to −3.29%. Finally, by applying optimum targeting, the treatment effects can be improved to −5.61%. This improvement is attributed to the assignment of households whose predicted treatment effects are positive if assigned to the nudge group to the rebate group and to the assignment of households whose predicted treatment effects are larger for the nudge intervention to the nudge group. The optimum assignment allocates 73.6% of households to the rebate intervention and the rest to the nudge intervention.

A question arises, then, regarding how a policymaker should implement targeting using observational household characteristics. In Table 6, we compare mean household characteristics over groups assigned to a treatment based on optimum targeting. Panels A and B of Table 6 present characteristics that are statistically significant at least at the 10% level.[29] Panel A includes electricity-usage-related characteristics, while Panel B relates to other household characteristics. Panel C lists the key variables presented in Figure 5, the mean for which is statistically indistinguishable among groups.

<Table 6 Comparison of household characteristics by targeting>

The results presented in Panel A suggest that households whose mean electricity usage during the pre-event period is relatively high or whose consumption in early winter is larger than that of similar households should be assigned to the rebate intervention rather than the nudge intervention. This result is also implied by Panel B of Figures 6 and 7, which shows that the nudge intervention does not encourage energy saving for households with high mean electricity usage in the pre-event period or larger consumption relative to similar households in early winter. In contrast, Panel A suggests that households whose mean electricity usage in the pre-event period is relatively low or whose consumption is less than that of similar households should be assigned to the nudge intervention rather than the rebate intervention. Panel B of Figures 6 and 7 also shows that the nudge intervention encourages energy saving for households with low mean electricity usage in the pre-event period or lower consumption relative to similar households. Thus, targeting based on electricity consumption can improve the treatment effects.[30]

The results in Panel A of Table 6 can be explained by other household characteristics listed in Panel B of Table 6. Compared to households assigned to the nudge group based on targeting, households in the rebate group tend to have older houses, larger household size, larger home size, double-story houses, and older heads of household. These characteristics are expected to be positively correlated with energy consumption. For example, the age of the house should be positively correlated with the age of durable electric goods, the home size should be proportional to the number of rooms, and the elderly probably stay home longer than the young. The larger the values of these characteristics, the larger the absolute values of the treatment effects, thus exhibiting the

[29] A complete table is reported in Appendix B.

[30] In Appendix C, we confirm the correlations between these electricity-usage-related characteristics and the treatment effect heterogeneity with the sample splitting approach proposed by Chernozhukov et al. (2019).

same association between electricity-usage-related characteristics and treatment effects.

The variables listed in Figure 5, the mean for which is statistically indistinguishable among the groups (i.e., variables listed in Panel C of Table 6), include the number of air conditioners and laptops, mean bills for electricity and gas, gender, income, and possession of solar panels. Although these characteristics are important for growing trees, the lack of statistical difference suggests that we cannot use such information for targeting.

The results suggest two approaches to targeting. The first approach is based on the simple mean energy consumption during the pre-event period or other observable household characteristics. The treatment effects can be improved, for example, by assigning the rebate intervention to households whose mean energy consumption is high or those who own an older or larger house and by assigning the nudge intervention to the remaining households. The second approach exploits information on the difference in electricity usage relative to similar households. The discussion so far reveals that such difference is determinant for explaining treatment effect heterogeneity (see Figure 7 and Table 6). We find similar households by applying RF based on electricity-usage-related variables and other household characteristics. This approach allows us to improve the treatment effects.

6. Conclusion

The findings of this study imply that the impacts of nonmonetary interventions can differ from those of monetary interventions in terms of heterogeneity. Hence, a policymaker using these interventions to internalize negative externalities as a public policy should not assess policy impacts based only on their mean effects. In addition, our results indicate that the impacts of monetary and nonmonetary interventions can be significantly improved by sophisticated targeting, which utilizes machine learning. The proposed targeting is practical, at least in the context of energy saving. A policymaker can efficiently manage demand by first predicting energy-saving behaviors during the high-demand season (e.g., mid-winter) with information on energy consumption during the preseason (e.g., early winter) and household characteristics and then assigning either monetary or nonmonetary interventions to each household based on these predictions.

The proposed targeting assigns an intervention to each household according to its predicted impacts. However, such differentiation faces a tradeoff between predicted improvement and costs (computation and information collection). The desirable degree of sophistication of the targeting procedure is an open issue. In addition, extensive differentiation in public policies might cause legal or ethical concerns (e.g., complaints against unequal or unfair interventions). A policymaker needs to discuss these issues carefully before implementing targeting and ensure the transparency of the differentiation process. Nonetheless, a policy design involving machine learning and targeting based on observable household characteristics to make maximum use of nonmonetary interventions may stand as a new form of policy mix.

We conclude this study by suggesting directions for future research. First, we separately evaluate the effects of monetary and nonmonetary interventions, and it is a natural extension of the current study to jointly measure them. The effects of the joint intervention might not be a simple summation of the effects of two separate interventions due to the crowding-in or -out effects of each intervention (Dolan and Metcalfe, 2015; Brandon et al., 2019). Specifically, treatment effect heterogeneity of the joint intervention can differ from the heterogeneities we detect in this study and would be of great interest to policymakers. Second, although we investigate the effects and heterogeneities of the short-period interventions, an understanding of those of the long-period interventions is valuable for policymakers to further improve policy impacts. As discussed in Section 5, energy-saving behaviors can differ if the interventions last for a longer period, possibly leading to different treatment effect heterogeneities. Therefore, future work based on long-period interventions could add interesting findings.

Acknowledgments

This work was supported by the Strategic Basic Research Programs of Core Research for Evolutional Science and Technology “Behavioral Economics Approach toward the Demand Response of Energy Consumption and the Social Development of Smart Grid” funded by the Japan Science and Technology Agency (JST). Kayo Murakami and Hideki Shimada equally contribute to this work. Please send inquiries to either Kayo Murakami or Takanori Ida.

References

[1] Allcott, H. (2011). Social norms and energy conservation", Journal of Public Economics 95(9), 1082-1095.

[2] Allcott H, Kessler JB, (2019). The welfare effects of nudges: A case study of energy use social comparisons. Am Econ J Appl Econ. Vol. 11(1): 236-276.

[3] Allcott, H. and S. Mullainathan. (2010). Behavior and Energy Policy. Science 327(5970):1204-1205.

[4] Allcott, H. and T. Rogers, (2014). The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation. American Economic Review 104(10): 3003-3037.

[5] Arimura, T., H. Katayama, and M. Sakudo (2016). Do Social Norms Matter to Energy-Saving Behavior? Endogenous Social and Correlated Effects, Journal of the Association of Environmental and Resource Economists 3(3): 525-553.

[6] Athey S, and GW. Imbens, (2016). Recursive partitioning for heterogeneous causal effects. PNAS, 113:7353–60.

[7] Athey S, and GW. Imbens, (2017). The state of applied econometrics: causality and policy evaluation. Journal of Economics Perspective, 31:3–32.

[8] Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, *47*(2), 1148-1178.

[9] Athey, S., and Wager, S. (2019). Estimating treatment effects with causal forests: An application. *arXiv preprint arXiv:1902.07409*.

[10] Ayres, I., S. Raseman, A. Shih, (2012). 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.

[11] Azarova, V., Jed J. Cohen, A. Kollmann, J. Reichl (2020). Reducing household electricity consumption during evening peak demand times: Evidence from a field experiment, Energy Policy 144, 111657.

[12] Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, *25*(2), 197-227.

[13] Borenstein S., (2002). The trouble with electricity markets: Understanding California's restructuring disaster. J Econ Perspect 16:191–211.

[14] Borenstein S., (2005). The long-run efficiency of real-time electricity pricing. Energy Journal 26: 93–116.

[15] Brandon, Alec, John A. List, Robert D. Metcalfe, Michael K. Price, Florian Rundhammer, (2018). 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; DOI: 10.1073/pnas.1802874115.
[16] Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.
[17] Bunk, A. P. and F. X. Gibbons, (2007). Social comparison: The end of a theory and the emergence of a field. Organizational Behavior and Human Decision Processes 102: 3-21.
[18] Carter, M. R., Tjernström, E., & Toledo, P. (2019). Heterogeneous impact dynamics of a rural business development program in Nicaragua. *Journal of Development Economics*, *138*, 77-98.
[19] Chernozhukov, V., Demirer, M., Duflo, E., & Fernandez-Val, I. (2019). *Generic machine learning inference on heterogenous treatment effects in randomized experiments.* Working Paper.
[20] Costa, D.L., Kahn, M.E., (2013). Energy conservation "nudge" and environmentalist ideology: evidence from a randomized residential field experiment. J. Eur. Econ. Assoc. 11, 680–702.
[21] Davis, J., & Heller, S. B. (2017a). Using causal forests to predict treatment heterogeneity: An application to summer jobs. *American Economic Review*, *107*(5), 546-50.
[22] Davis, J., & Heller, S. B. (2017b). Rethinking the benefits of youth employment programs: The heterogeneous effects of summer jobs. *Review of Economics and Statistics*, 1-47.
[23] Dolan, P., & Metcalfe, R. (2015). Neighbors, knowledge, and nuggets: two natural field experiments on the role of incentives on energy conservation. *Becker Friedman Institute for Research in Economics Working Paper*, (2589269).
[24] Faruqui A, Sergici S (2010) Household response to dynamic pricing of electricity: A survey of 15 experiments. J Regul Econ 38:193–225. 22.
[25] Festinger, L. (1954). A theory of social comparison process. Human Relations 1: 117-140.
[26] Fowlie, Meredith, Catherine Wolfram, C. Anna Spurlock, Annika Todd, Patrick Baylis, and Peter Cappers (2020) Default Effects and Follow-on Behavior: Evidence from an Electricity Pricing Program. URL: https://www.meredithfowlie.com/research
[27] Gillan, James M. (2017) Dynamic Pricing, Attention, and Automation: Evidence from a Field Experiment in Electricity Consumption. Haas WP 284.
[28] Glaeser, Edward L. (2006). "Paternalism and Psychology." University of Chicago Law Review 73 (1): 133–56.
[29] Goldstein, J., B., Cialdini, and V., Griskevicius, (2008). A room with a viewpoint:

Using social norms to motivate environmental conservation in hotels. Journal of consumer Research 35(3): 472-482.

[30] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

[31] Ida, T., Murakami, K., & Tanaka, M. (2016). Electricity demand response in Japan: Experimental evidence from a residential photovoltaic power-generation system. *Economics of Energy & Environmental Policy*, *5*(1), 73-88.

[32] Ito, K., T. Ida, and M. Tanaka (2018). Moral Suasion and Economic Incentives: Field Experimental Evidence from Energy Demand. American Economic Journal: Economic Policy 10(1), 240-267.

[33] Ito, K., T. Ida, and M. Tanaka. (2017). Information Frictions, Inertia, and Selection on Elasticity: A Field Experiment on Electricity Tariff Choice. Working Paper.

[34] Joskow PL., (2012). Creating a smarter U.S. electricity grid. J Econ Perspect 26:29–48.

[35] Knittel, C. R., & Stolper, S. (2019). *Using Machine Learning to Target Treatment: The Case of Household Energy Use* (No. w26531). National Bureau of Economic Research.

[36] Kurz, T., N. Donaghue, and I. Walker (2005). Utilizing a Social-Ecological Framework to Promote Water and Energy Conservation: A Field Experiment. Journal of Applied Social Psychology 35(6): 1281-1300.

[37] Martin, S. and N. Rivers (2018). Information Provision, Market Incentives, and Household Electricity Consumption: Evidence from a Large-Scale Field Deployment," Journal of the Association of Environmental and Resource Economists 5(1): 207-231.

[38] Nolan, J. W., P. W. Schultz, R. B. Cialdini, N. J. Goldstein, and V. Griskevicius, (2008). Normative Social Influence is Underdetected. Personality & social psychology bulletin 34(7): 913-923.

[39] O'Neill, E., & Weeks, M. (2018). Causal Tree Estimation of Heterogeneous Household Response to Time-Of-Use Electricity Pricing Schemes. *arXiv preprint arXiv:1810.09179*.

[40] Schultz, P. W., (1999). Changing behavior with normative feedback interventions: a field experiment on curbside recycling. Basic Appl. Soc. Psychol. 21: 25-36.

[41] Schultz, P. W., J. M. Nolan, R. B. Cialdini, N. J. Goldstein, V. Griskevicius. (2007). The Constructive, Destructive, and Reconstructive Power of Social Norms. Psychological Science 18(5): 429-34.

[42] Tibshirani, J., Athey S., Wager S., Friedberg R., Miner L., & Wright M. (2018). *Package 'grf'*.

[43] Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228-1242.

[44] Wang, X. and W. Tang, "To Overconsume or Underconsume: Baseline Manipulation in Demand Response Programs," 2018 North American Power Symposium (NAPS), Fargo, ND, 2018, pp. 1-6, doi: 10.1109/NAPS.2018.8600558.

[45] Wolak, F. A. (2007). Residential Customer Response to Real-time Pricing: The Anaheim Critical Peak Pricing Experiment. UC Berkeley: Center for the Study of Energy Markets. Retrieved from https://escholarship.org/uc/item/3td3n1x1

[46] Wolak FA (2011) Do residential customers respond to hourly prices? Evidence from a dynamic pricing experiment. Am Econ Rev 101:83–87. 21.

Figure 1. Timeline and procedures of the experiment

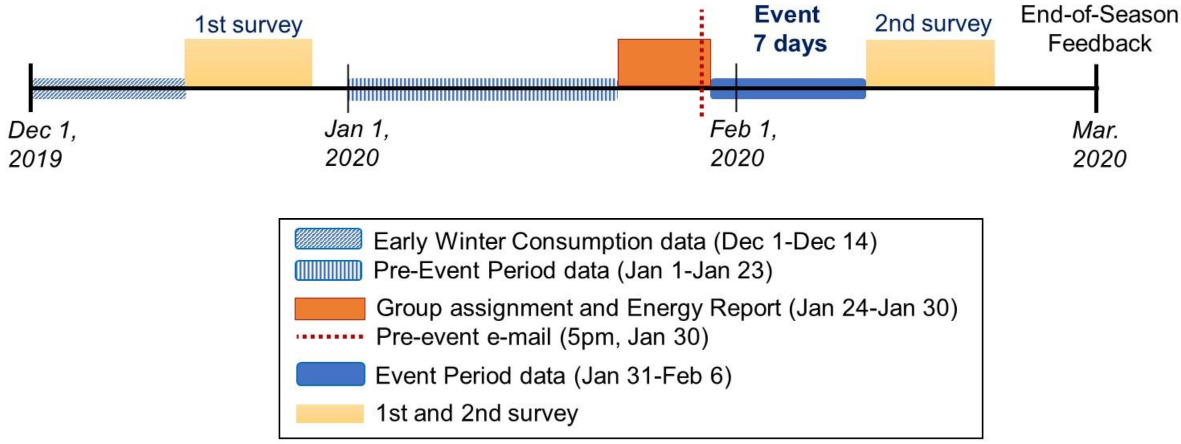
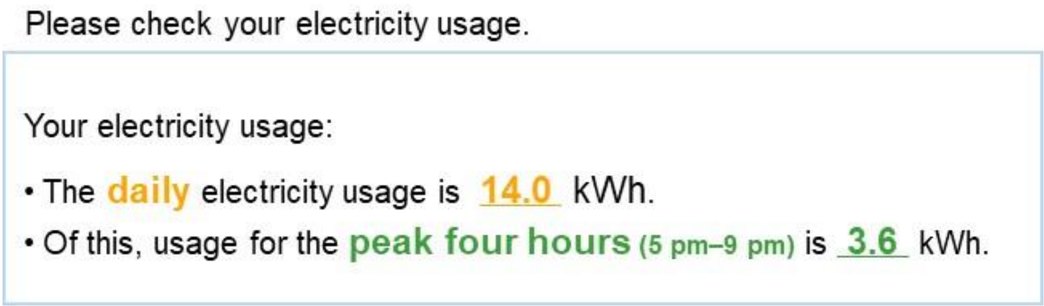


Figure 2. Energy report for the control and rebate groups

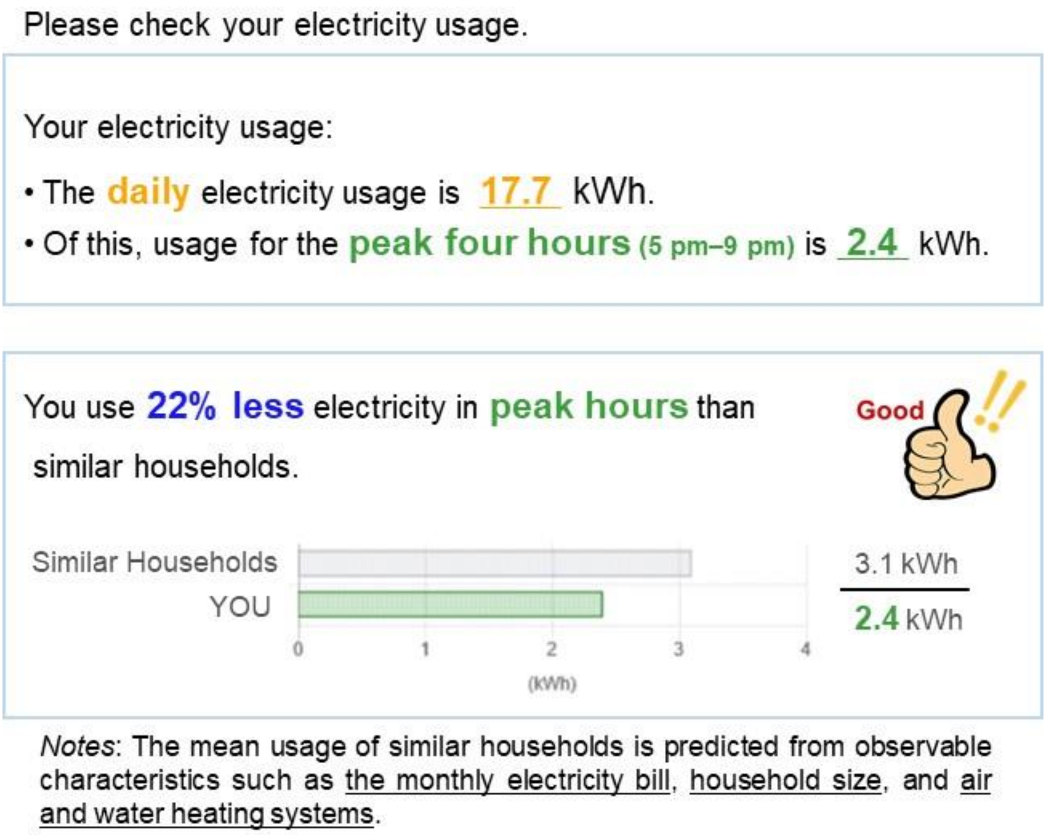

Please check your electricity usage.

Your electricity usage:

- The daily electricity usage is 14.0 kWh.
- Of this, usage for the peak four hours (5 pm–9 pm) is 3.6 kWh.

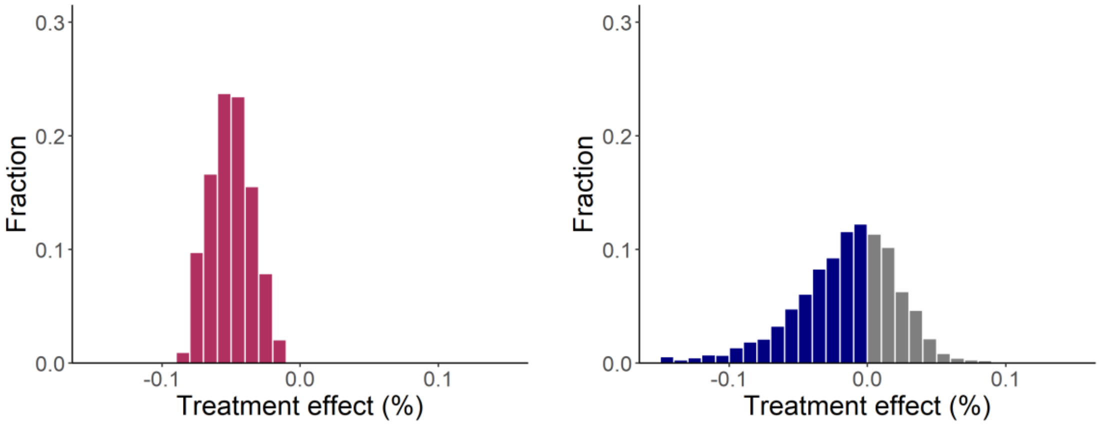
Note: This figure is translated into English from the Japanese energy report.

Figure 3. The Energy Report for the nudge group



Note: This figure is translated into English from the Japanese energy report.

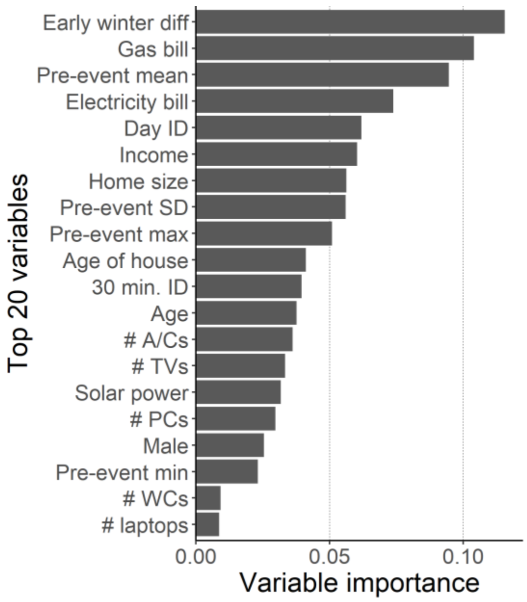
Figure 4. Distributions of heterogenous treatment effects



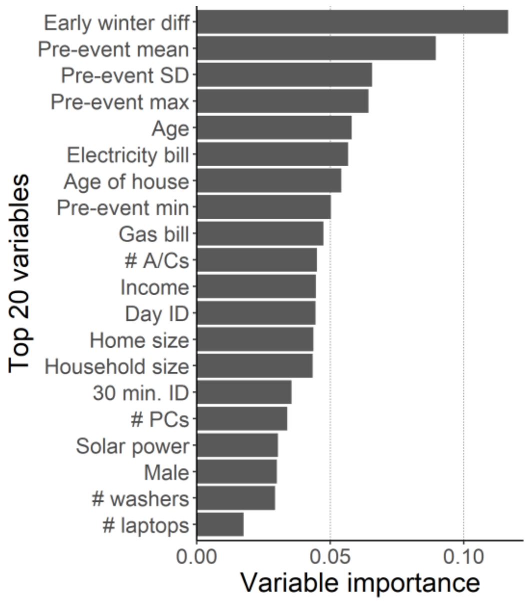
(A) Rebate

(B) Nudge

Figure 5. Key variables for growing trees



(A) Rebate



(B) Nudge

Figure 6. Mean electricity usage in the pre-event period and treatment effects

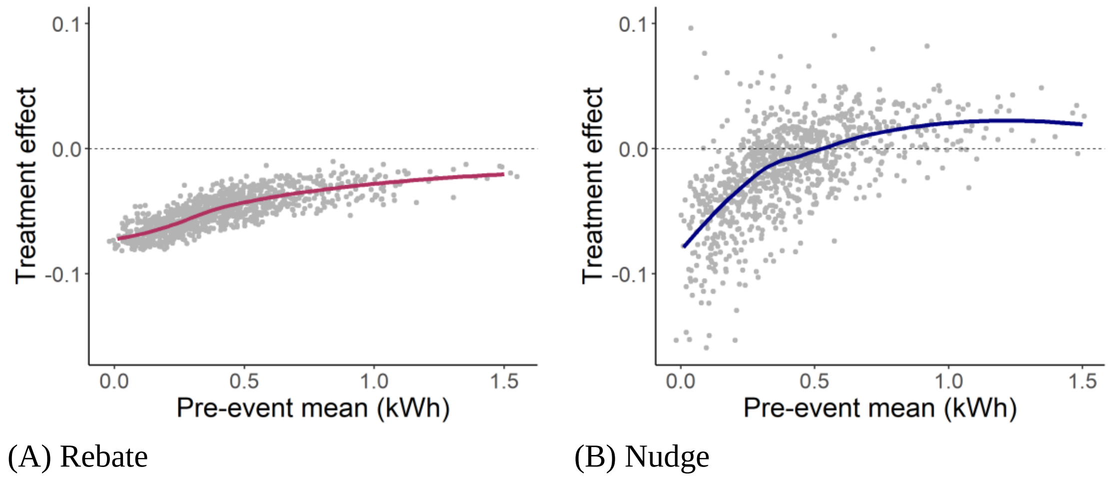
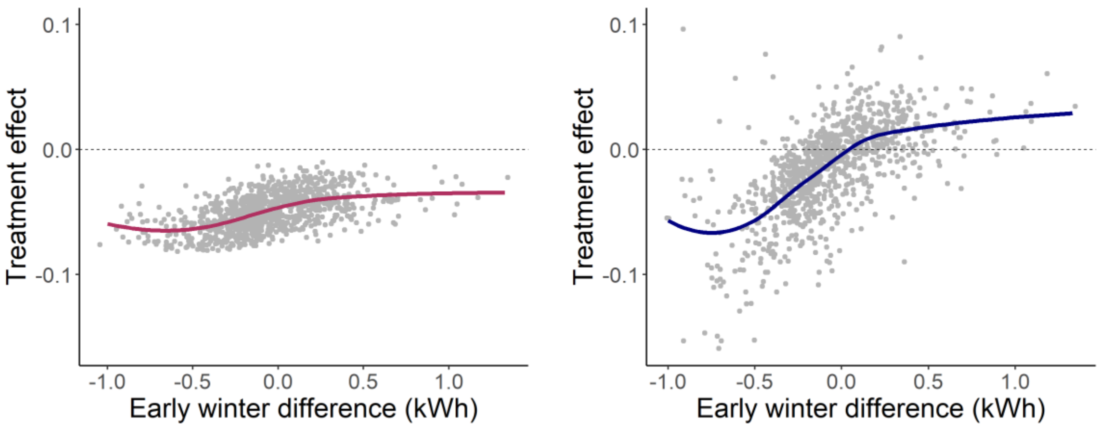


Figure 7. Difference in electricity usage and treatment effects



(A) Rebate (B) Nudge

Table 1. Summary statistics by control group and treatment groups

	Control (N=327)	Rebate (N=313)		Nudge (N=314)	
	Average	Difference	p-value	Difference	p-value
Early Winter (December 1–14)					
Electricity use (kWh /day)	13.818	0.084	0.909	−0.114	0.877
Electricity use (kWh /peak-time)	2.761	0.010	0.941	0.006	0.967
Pre-Event Period (January 1–23)					
Electricity use (kWh /day)	16.876	0.091	0.920	0.009	0.992
Electricity use (kWh /peak-time)	3.343	−0.009	0.957	0.051	0.770
Rebate Baseline (January 17–23)					
Electricity use (kWh /day)	17.029	0.082	0.929	0.018	0.984
Electricity use (kWh /peak-time)	3.383	−0.054	0.750	0.072	0.688
Demographic Characteristics					
Household size (persons)	2.700	0.012	0.898	−0.028	0.772
Number of A/Cs	2.994	0.143	0.294	0.025	0.855
Home size (Square meter)	116.667	−2.242	0.551	−0.297	0.938
Household income (JPY/million)	6.343	−0.366	0.232	−0.070	0.828
All electric house (Dummy)	0.440	−0.054	0.168	−0.039	0.317

Note: The winter consumption pattern of electricity use begins to be observed in the "Early Winter" season in Japan. We use this period's data for calculating social comparison information in the energy report for households in the nudge group. "Pre-Event Period" data are used for baseline consumption to estimate the ATE. "Rebate Baseline" data are used for calculating rewards for energy conservation for households in the rebate group, who are informed regarding the rebate calculation rule after January 24.

Table 2. ATEs of rebate and nudge

	All household		Subgroup			
	N=954		Less (N=581)		More (N=355)	
Rebate	−0.043	***	−0.056	***	−0.025	
	(0.013)		(0.016)		(0.023)	
Nudge	−0.007		−0.038	**	0.036	
	(0.013)		(0.015)		(0.024)	
Observations	214,173		126,598		83,372	

Note: **p<0.05, ***p<0.01. This table shows the DID estimation results for Equation (2). Standard errors are reported in parentheses. We used the natural logarithm of electricity usage for the dependent variable; hence, the treatment effects may be approximately interpreted in percentage terms.

Table 3. ATEs computed by the GRF algorithm

	Rebate		Nudge
Average treatment effects	−0.052	***	−0.014
	(0.016)		(0.015)

Note: ***p<0.01. The number of observations is 224,400. Due to the normalization by households' mean consumption during the pre-event period, the results in this table measure percent point changes. Standard errors are reported in parentheses.

Table 4. Results of the slope test

	ATE (γ)	Heterogeneity (η)
Rebate	−0.044	1.641
	[−0.073, −0.015]	[0.050, 3.296]
Nudge	−0.008	1.410
	[−0.032, 0.016]	[0.051, 2.793]

Note: The numbers in brackets represent the 90% confidence interval proposed by Chernozhukov et al. (2019). The number of iterations is 1000.

Table 5. Improvement in treatment effects by targeting

	Net treatment effects	
	Uniform	Targeting
Rebate	−5.02%	−5.02%
	(1.49)	(1.49)
Nudge	−1.25%	−3.29%
	(5.77)	(3.60)
Optimum targeting	−5.61%	
	(2.33)	

Note: "Uniform" measures the treatment effects predicted under the scenario in which all households are uniformly assigned to one intervention. "Targeting" measures the treatment effects that are realized if an intervention is assigned only to households with negative treatment effects. "Optimum Targeting" represents the treatment effects realized if an intervention with larger effects is assigned.

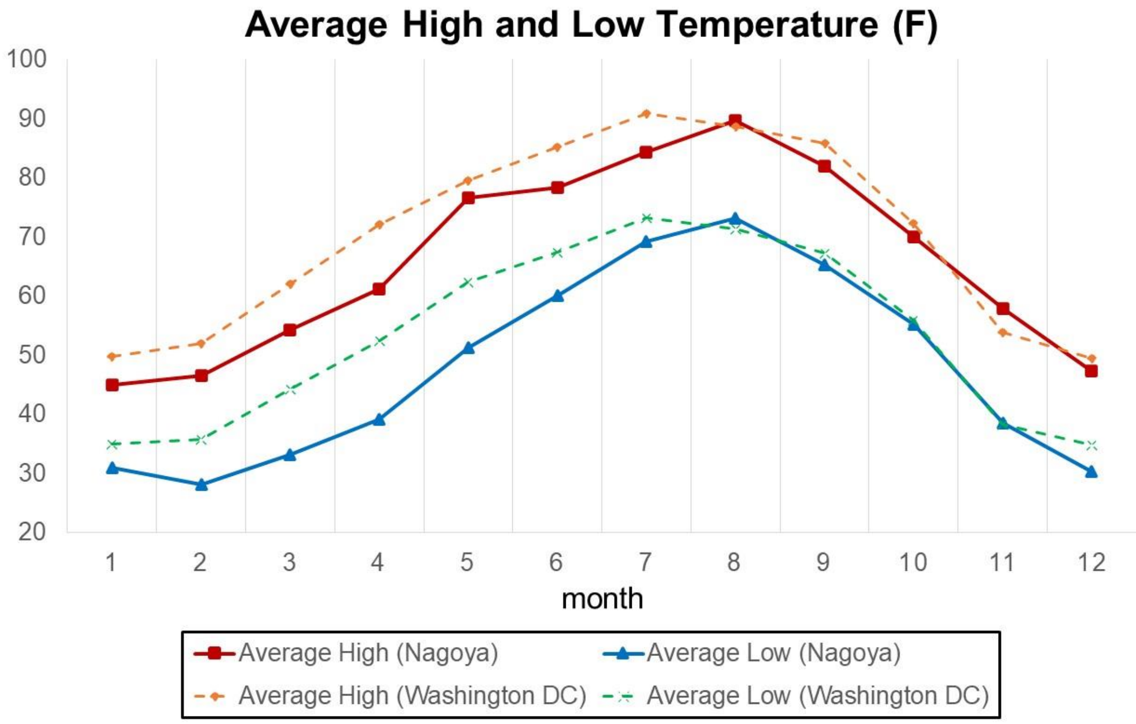
Table 6. Comparison of household characteristics by targeting

	Rebate (*N* = 688)	Nudge (*N* = 247)	Difference	S.E.	p-value
A. *Energy Consumption*					
Pre-event mean (kWh)	0.469	0.292	0.178	0.016	0.000
Early winter difference (kWh)	0.022	−0.300	0.322	0.018	0.000
B. *Significant Variables for Targeting*					
Age of house (years)	22.093	15.320	6.773	0.703	0.000
Household size (persons)	2.756	2.583	0.173	0.089	0.053
Home size (m^2)	120.051	108.806	11.245	3.612	0.002
Double-story (Dummy)	0.712	0.599	0.113	0.036	0.002
Age (years/10)	5.269	4.676	0.593	0.089	0.000
C. *Insignificant Variables Listed in Figure 5*					
Num. of A/Cs	3.129	2.951	0.178	0.134	0.186
Num of laptops	1.391	1.336	0.055	0.078	0.483
Electricity bill (JPY/10,000)	0.971	0.899	0.072	0.063	0.260
Gas bill (JPY/10,000)	0.319	0.285	0.034	0.025	0.166
Male (Dummy)	0.783	0.769	0.014	0.031	0.648
Income (JPY/10 million)	0.629	0.608	0.021	0.026	0.435
Solar power (Dummy)	0.246	0.243	0.003	0.032	0.932

Note: The variables in Panel B exhibit statistically significant differences up to the 10% level among groups. A complete table is presented in Appendix B.

Appendix A: Additional information

Figure A1. The latest average high and low temperatures in Nagoya, Japan and Washington DC, United States



Note: This figure shows the latest temperatures (F) from April 2019 to March 2020, obtained from the "Tokyo Climate Center, WMO Regional Climate Center, World Climate."
https://ds.data.jma.go.jp/gmd/tcc/tcc/products/climate/index.html

Figure A2. The messages and rebate rule information for the rebate group

Substantial energy conservation will be required for the society during the coldest part of winter (January and February in particular).

Please reduce your electricity usage in the critical peak-demand hours during the following week.

IMPORTANT NOTICE

- During **5 pm – 9 pm** (critical peak-demand hours) of the event week from **January 31 to February 6**, you will receive **rebates** for your electricity conservation in addition to **your participation rewards**.
- Rebate will be **100 yen per kWh** (**maximum 1,000 yen** in total per week).
- Your total electricity conservation will be calculated as the reduction from your usage during the normal week of January 17 - 23.

Note: This figure reports the message regarding the rebate calculation rule for the households in the rebate group (R), translated into English from the original Japanese version. The households received this message sometime between January 24 and January 30.

Appendix B. Additional tables

Table B1. Comparison of mean household characteristics across groups assigned by targeting

	Rebate (N = 688)	Nudge (N = 247)	Difference	S.E.	p-value
	A. *Baseline Variable*				
Pre-event mean (kWh)	0.469	0.292	0.178	0.016	0.000
Pre-event max (kWh)	1.191	0.949	0.242	0.043	0.000
Pre-event min (kWh)	0.103	0.038	0.065	0.006	0.000
Pre-event S.D. (kWh)	0.215	0.176	0.039	0.008	0.000
Difference in early winter (kWh)	0.022	−0.300	0.322	0.018	0.000
	B. *Significant Variables for Targeting*				
Age of house (years)	22.093	15.320	6.773	0.703	0.000
Age (years/10)	5.269	4.676	0.593	0.089	0.000
TOU (dummy)	0.411	0.482	−0.070	0.037	0.057
Double-story (dummy)	0.712	0.599	0.113	0.036	0.002
Home size (m^2)	120.051	108.806	11.245	3.612	0.002
HW: Eco-cute (dummy)	0.326	0.441	−0.116	0.036	0.002
HW: Eco-jaws (dummy)	0.131	0.085	0.046	0.022	0.037
IH heater (dummy)	0.390	0.457	−0.068	0.037	0.066
Accumulator (dummy)	0.026	0.057	−0.031	0.016	0.057
Num. of PCs	2.109	1.883	0.226	0.092	0.014
Num. of desktop PCs	0.718	0.547	0.171	0.057	0.003
Num. of driers	0.346	0.279	0.067	0.034	0.052
Num. of refrigerators	1.286	1.190	0.096	0.041	0.019
Num. of gas ovens	0.109	0.053	0.056	0.019	0.003
Num. of TVs	2.314	2.016	0.298	0.095	0.002
Num. of solar calorifiers	0.041	0.012	0.029	0.010	0.006
Num. of heaters	1.016	0.810	0.206	0.063	0.001
Num. of pets	0.814	0.862	−0.048	0.027	0.069
Attitude (1, 2, …,5)	0.802	0.858	−0.056	0.027	0.038
Job: employee (dummy)	0.404	0.571	−0.167	0.037	0.000
No job (dummy)	0.321	0.211	0.111	0.032	0.000
Educ 1 (dummy)	0.311	0.231	0.080	0.032	0.013
Household size (persons)	2.756	2.583	0.173	0.089	0.053
	C. *Insignificant Variables Listed in Figure 5*				
Electricity bill (JPY/10,000)	0.971	0.899	0.072	0.063	0.260
Gas bill (JPY/10,000)	0.319	0.285	0.034	0.025	0.166
Num. of A/Cs	3.129	2.951	0.178	0.134	0.186
Num. of laptops	1.391	1.336	0.055	0.078	0.483
Male (dummy)	0.783	0.769	0.014	0.031	0.648
Income (JPY/10 million)	0.629	0.608	0.021	0.026	0.435
Solar power (dummy)	0.246	0.243	0.003	0.032	0.932

Table B2. Comparison of the mean electricity usage and mean household characteristics across subgroups

	Less (N=581)	More (N=355)	Difference	*p*-value
Early Winter				
Daily electricity use (kWh)	11.013	18.498	−7.485	0.000
Peak-time electricity use (kWh)	1.918	4.167	−2.248	0.000
Pre-Event Period				
Daily electricity use (kWh)	13.914	21.939	−8.025	0.000
Peak-time electricity use (kWh)	2.454	4.848	−2.394	0.000
Rebate Baseline				
Daily electricity use (kWh)	14.025	22.186	−8.161	0.000
Peak-time electricity use (kWh)	2.471	4.911	−2.439	0.000
Demographic characteristics				
Household size (persons)	2.570	2.927	−0.357	0.000
Num. of A/Cs	2.966	3.203	−0.237	0.045
Home size (m^2)	112.806	121.507	−8.702	0.006
Income (JPY/ million)	6.060	6.500	−0.440	0.095
All electric house (dummy)	0.386	0.448	−0.062	0.060

Table B3. Mean household characteristics at various quartiles of the treatment effects

	Rebate				Nudge			
	1st	2nd	3rd	4th	1st	2nd	3rd	4th
Treatment Effects	−0.069	−0.056	−0.045	−0.031	−0.064	−0.023	−0.001	0.027
Energy Consumption								
Pre-event mean (kWh)	0.184	0.323	0.484	0.699	0.210	0.358	0.489	0.633
Pre-event max (kWh)	0.602	0.914	1.275	1.719	0.708	1.020	1.232	1.549
Pre-event min (kWh)	0.018	0.052	0.094	0.179	0.016	0.061	0.116	0.150
Pre-event S.D. (kWh)	0.114	0.165	0.230	0.310	0.136	0.185	0.221	0.277
Early winter difference (kWh)	−0.283	−0.116	0.008	0.140	−0.337	−0.141	0.033	0.194
Other Household Characteristics								
Age of house (years)	21.194	18.641	21.122	20.258	17.400	19.530	22.158	22.135
Male (dummy)	0.645	0.744	0.838	0.893	0.679	0.769	0.778	0.893
Age (years/10)	4.838	4.910	5.286	5.416	4.650	5.090	5.265	5.446
Electricity bill (JPY/10,000)	0.553	0.805	1.070	1.382	0.698	0.895	1.042	1.174
Gas bill (JPY/10,000)	0.238	0.368	0.331	0.303	0.309	0.290	0.325	0.317
Home size (m^2)	0.887	1.016	1.232	1.551	0.976	1.121	1.270	1.317
Solar power (dummy)	0.115	0.205	0.248	0.412	0.171	0.248	0.214	0.348
Num. of A/Cs	2.115	2.684	3.269	4.266	2.453	2.970	3.372	3.536
Num. of PCs	1.714	1.897	2.060	2.528	1.731	1.957	2.205	2.305
Num. of laptops	1.175	1.406	1.325	1.601	1.261	1.261	1.453	1.532
Num. of TVs	1.641	1.957	2.316	3.030	-	-	-	-
Num. of WCs	0.927	1.188	1.415	1.704	-	-	-	-
Num. of washers	-	-	-	-	0.457	0.568	0.585	0.438
Income (JPY/10 million)	0.492	0.596	0.625	0.781	0.541	0.591	0.659	0.702
Household size (persons)	-	-	-	-	2.274	2.577	2.902	3.090

Appendix C. Association between heterogeneous treatment effects and household characteristics

Table C1. Classification analysis

	Mean electricity use in the pre-event period (kWh)			Difference from similar households (kWh)		
	Q1	Q4	Q1-Q4	Q1	Q4	Q1-Q4
Rebate	0.223	0.662	−0.441	−0.243	0.118	−0.362
	[0.216, 0.231]	[0.645, 0.678]	[−0.459, −0.424]	[−0.259, −0.227]	[0.101, 0.134]	[−0.385, −0.338]
Nudge	0.197	0.667	−0.471	−0.347	0.202	−0.549
	[0.190, 0.204]	[0.647, 0.686]	[−0.492, −0.450]	[−0.360, −0.333]	[0.185, 0.219]	[−0.571, −0.528]

Note 1: The sample splitting approach proposed by Chernozhukov et al. (2019) allows us to examine the relationship between the treatment effects and various covariates, which is called classification analysis (CLAN). Households are divided into subgroups based on the size of the predicted treatment effects. Then, the means of the covariates of interest are compared across groups. The CLAN allows researchers to find the types of households that are most affected by a treatment. In this study, we focus on mean electricity use during the pre-event period and the difference in electricity usage relative to similar households in early winter in Table C1. To calculate the values in Table C1, we divide households based on quartiles and calculate the means of the characteristics for households in the top (Q1) and bottom (Q4) quartiles.

Note 2: The numbers in brackets represent the 90% confidence interval proposed by Chernozhukov et al. (2019). The number of iterations is 1000. Table C1 shows the mean electricity use during the pre-event period. The difference in electricity usage relative to similar households is statistically different among groups. These results support the validity of targeting, as discussed in Section 5.