NBER WORKING PAPER SERIES

# DYNAMIC TARGETING: EXPERIMENTAL EVIDENCE FROM ENERGY REBATE PROGRAMS

Takanori Ida
Takunori Ishihara
Koichiro Ito
Daido Kido
Toru Kitagawa
Shosei Sakaguchi
Shusaku Sasaki

Working Paper 32561
http://www.nber.org/papers/w32561

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2024, Revised April 2026

We would like to thank Linnea Holy and Tomoki Nishiyama for their exceptional research assistance and Christopher Costello, Michael Greenstone, Ryan Kellogg, Anna Russo, Yuta Toyama, and seminar participants at the Coase Project Conference, Bologna, Cornell, DUFE, HKU, 2024 LAMES, LMU Munich, Northwestern, and UC Berkeley Energy Camp for their helpful comments. We thank the Ministry of the Environment, Government of Japan (MOEJ) for their collaboration in realizing this study. Kitagawa and Sakaguchi gratefully acknowledge financial support from the ERC grant (number 715940) and the ESRC Centre for Microdata Methods and Practice (CeMMAP) (grant number RES-589-28-0001). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Dynamic Targeting: Experimental Evidence from Energy Rebate Programs
Takanori Ida, Takunori Ishihara, Koichiro Ito, Daido Kido, Toru Kitagawa, Shosei Sakaguchi, and Shusaku Sasaki
NBER Working Paper No. 32561
June 2024, Revised April 2026
JEL No. C1, C50, C9, C93, Q4, Q40, Q41, Q48, Q50, Q58

**ABSTRACT**

Economic policies often involve dynamic interventions, in which individuals receive repeated treatments over multiple periods. However, existing economic studies typically focus on static targeting, potentially overlooking valuable information generated by dynamic interventions. We develop a framework for designing optimal dynamic targeting policies that maximize social welfare. We show that dynamic targeting can outperform static targeting through several key mechanisms—learning, habit formation, and adaptive targeting effects—and develop a method to empirically identify these effects. Using a sequential randomized controlled trial (RCT) of a residential energy rebate program, we find that dynamic targeting nearly doubles the social welfare gains relative to static targeting and produces more than five times larger welfare gains than non-targeting policies. We also investigate how households' anticipatory responses to adaptive assignment rules affect the welfare gains from dynamic targeting.

Takanori Ida
Kyoto University
ida@econ.kyoto-u.ac.jp

Takunori Ishihara
Kyoto University of Advanced Science
ishihara.takunori@kuas.ac.jp

Koichiro Ito
University of Chicago
Harris School of Public Policy
and NBER
ito@uchicago.edu

Daido Kido
Otaru University of Commerce
dkido@res.otaru-uc.ac.jp

Toru Kitagawa
Brown University
Department of Economics
and Institute For Fiscal Studies - IFS
toru_kitagawa@brown.edu

Shosei Sakaguchi
The University of Tokyo
sakaguchi@e.u-tokyo.ac.jp

Shusaku Sasaki
Osaka University
ssasaki.econ@gmail.com

# 1 Introduction

Targeting has become a central question in economics and policy design. When policymakers face budget constraints, identifying who should be treated is critical for enhancing policy effectiveness across various policy domains, including job training and mentoring programs (Kitagawa and Tetenov, 2018; Athey and Palikot, 2026), social safety net programs (Finkelstein and Notowidigdo, 2019; Deshpande and Li, 2019), energy efficiency programs (Burlig, Knittel, Rapson, Reguant, and Wolfram, 2020), electricity conservation nudges (Knittel and Stolper, 2021), dynamic electricity pricing (Ito, Ida, and Tanaka, 2023), and energy rebate programs (Ida, Ishihara, Ito, Kido, Kitagawa, Sakaguchi, and Sasaki, 2026).

To date, the economic literature has primarily focused on *static* targeting for one-time policy interventions. However, many real-world economic policies involve *dynamic* interventions. Such policies—such as job training programs (Lechner, 2009; Rodríguez, Saltiel, and Urzúa, 2022), unemployment insurance programs (Meyer, 1995; Kolsrud, Landais, Nilsson, and Spinnewijn, 2018), healthcare programs (Luckett, Laber, Kahkoska, Maahs, Mayer-Davis, and Kosorok, 2019), and educational interventions (Ding and Lehrer, 2010)—administer repeated treatments over multiple periods. Individuals' responses to these interventions can vary with the timing and frequency of treatment, and data from earlier periods could provide valuable information to identify optimal subsequent assignment. By leveraging these dynamics and the resulting heterogeneity in responses, dynamic targeting has the potential to outperform static targeting. It is therefore crucial to study the design of optimal dynamic targeting and to understand the mechanisms through which it can achieve superior welfare outcomes over static targeting, which may overlook key information.

In this paper, we demonstrate how researchers can use experimental or quasi-experimental data to design dynamic targeting policies that maximize the social welfare gains from policy interventions. Our framework identifies the variation in the data required to estimate dynamic targeting policies and the mechanisms through which dynamic targeting enhances welfare. Applying the framework to a sequential randomized controlled trial (RCT) of an energy rebate program, we empirically show how dynamic targeting can enhance welfare gains relative to conventional static targeting and non-targeted policies.

Our dynamic targeting framework consists of a sequence of individualized treatment allocation policies that use two types of information to optimize treatment assignment. The first type draws on pre-intervention data, including household demographics, housing characteristics, and historical electricity usage. The second type incorporates the history of past interventions and each household's responses to them. Using the cumulative information available in each period, dynamic targeting determines the treatment to be assigned to each household.

Our theory derives four key mechanisms through which dynamic targeting can enhance social welfare. The first is the set of *period-specific treatment effects*, which capture the direct contemporaneous treatment effect within each period, separated from the dynamic effects arising from the other mechanisms described below.

The second is the *learning and habituation effect*. Individuals who receive an intervention over multiple periods may learn to respond more effectively to the treatment (for example, learning to conserve electricity in response to a rebate incentive). For individuals with a pronounced learning effect, repeated exposure to the treatment can yield greater welfare gains than a single, early-period intervention. Conversely, individuals who receive an intervention over multiple periods may become habituated to it, resulting in a diminished treatment response in later periods (Ito, Ida, and Tanaka, 2018).

The third is the *habit formation effect*. Individuals may develop lasting habits—such as conserving electricity—when repeatedly exposed to a program (Becker and Murphy, 1988). Consequently, even after leaving the program, they may continue to exhibit lower electricity consumption than those who were never treated. For individuals with strong habit formation, early interventions can yield persistent welfare gains even in the absence of subsequent treatments.

The fourth is the *adaptive targeting effect*. Dynamic targeting enables policymakers to gather information about individuals' responses after each intervention, which can be used to improve welfare through adaptive assignments of the future treatments. For example, consider a two-period binary treatment setting. In the first period, we could assign an individual to either treatment or no-treatment. For some individuals, observing the response to treatment in the first period is more informative to predict treatment responses in the second period and help the planner determine their treatment assignment for the second period; for others, the response to no-treatment may reveal more useful information for the planner's treatment assignment in the second period. In some cases, a statically suboptimal first-period treatment assignment may nonetheless be dynamically optimal if the information generated through dynamic information acquisition substantially improves targeting decisions in subsequent periods.

Our theory also clarifies what variation in the data is required to identify the optimal dynamic policy assignment. With this insight, we designed a randomized field experiment on a residential electricity rebate program in collaboration with the Ministry of the Environment, Government of Japan (MOEJ). The rebate program aims to encourage energy conservation during peak demand hours when the marginal cost of electricity is significantly higher than the time-invariant residential electricity price. In our context, the social welfare gain from this program could vary across individuals and can be positive, negative, or zero, given the per-household implementation cost, which suggests that optimal targeting could enhance the social welfare gain from this program.

Our sequential randomization consists of a two-period experiment, randomly assigning customers to one of four groups: $(U, U)$, $(T, T)$, $(U, T)$, and $(T, U)$. Here, $U$ denotes "untreated" and $T$ signifies "treated." Each element within the parentheses represents the treatment assignment for the first and second periods, respectively. We use data from this experiment with the Empirical Welfare Maximization (EWM) approach (Kitagawa and Tetenov, 2018; Sakaguchi, 2025), also referred to as Outcome Weighted Learning (Zhao, Zeng, Rush, and Kosorok, 2012), to estimate dynamic targeting.

We empirically compare the welfare benefits of four non-targeting policies, static targeting, and dynamic targeting. Our findings indicate that conventional static targeting significantly enhances welfare benefits relative to non-targeting policies. Moreover, dynamic targeting further augments these benefits, nearly doubling the welfare gain from static targeting policies.

We decompose the welfare gains into periods 1 and 2 to examine differences in welfare across the two periods. We find that the dynamic targeting policy generates *lower* welfare gains than the static targeting policy in period 1, but substantially higher welfare gains in period 2. This empirical finding is consistent with our theoretical prediction that a statically suboptimal first-period treatment assignment may nonetheless be dynamically optimal for maximizing total welfare gains across the two periods.

We then use our theory to empirically decompose the welfare gains from dynamic targeting into four mechanisms—period-specific treatment effects, learning, habit formation, and adaptive targeting effects. In particular, we find that adaptive targeting effects play a key role in enhancing welfare under dynamic targeting, and we identify the types of households that exhibit large adaptive targeting effects.

In our sequential RCT, households were informed that the treatments are randomized over the periods so that they had no incentive to manipulate their first-period responses to influence their second-period treatment assignment. When the planner implements a dynamic assignment policy, however, it is important to consider the possibility and welfare implications of households' anticipated responses in period 1 through the knowledge of, or anticipation of, the planner's adaptive assignment of the treatment in period 2. This can be viewed as an external validity issue for learning dynamic targeting with sequential RCT.

To address this external validity issue, we investigate how the welfare gains from the optimal targeting policy—which is optimal in the absence of anticipatory response—would change when households manipulate their first-period responses so as to receive their favored treatments in the second period with the knowledge of the targeting policy rule. Across different assumptions on the type of manipulation, we find that the welfare gains under the dynamic targeting policy remain close to the no-manipulation benchmark, with only modest differences. Our findings suggest that although the external validity issue by anticipatory responses is conceptually important, it is unlikely to materially change the magnitude of the welfare gain in our context.

*Related literature and our contributions*—Our study makes four primary contributions to the literature. First, we provide one of the first theoretical and empirical analyses of dynamic targeting in economics. As discussed earlier, existing economics studies have primarily focused on static targeting, but many real-world economic policies involve dynamic interventions. To our knowledge, our study is the first to develop a framework for dynamic targeting in economic policy and empirically test its predictions using a sequential RCT conducted in the field.[1]

Second, our theory uncovers the mechanisms underlying dynamic targeting. Proposition 2.1 decomposes the welfare gain from dynamic targeting into four key components—the period-specific treatment effects, the learning effect, the habit-formation effect, and the adaptive targeting effect. In the medical literature—particularly in the context of clinical trials—dynamic interventions have been studied under the framework of dynamic treatment regimes (DTRs), but the mechanisms underlying optimal dynamic interventions have not been formally analyzed. Our theory clarifies how dynamic targeting exploits information obtained from multi-period interventions.[2]

Third, we develop an econometric framework to estimate each mechanism described in Proposition 2.1. Specifically, we show in Proposition 5.1 how researchers can use experimental or quasi-experimental variation in treatment assignment to identify the four key components underlying dynamic targeting. We then empirically estimate these components using our experimental data. The resulting estimates illuminate how dynamic targeting leverages information from multi-period interventions to optimally predict and exploit the four key behavioral responses to treatment.[3]

Fourth, we develop a framework of robustness analysis against the external validity concern due to households' anticipatory responses. Our approach constructs the base-case and worst-case

[1]Related studies in the marketing literature include Ko, Uetake, Yata, and Okada (2022) and Liu (2023). Although their focus is not targeting in economic policy, they study dynamic coupon distribution in online retail settings using a Q-learning approach.

[2]Our framework builds on the dynamic treatment regime literature in medicine (Robins, 1986; Murphy, 2003; Chakraborty and Moodie, 2013; Zhang, Laber, Davidian, and Tsiatis, 2018; Tsiatis, Davidian, Holloway, and Laber, 2019). In medical science, determining the optimal individualized sequence and timing of treatments to improve health outcomes is often crucial (Pelham Jr, Fabiano, Waxmonsky, Greiner, Gnagy, Pelham III, Coxe, Verley, Bhatia, Hart et al., 2016). This parallels our research question in the social sciences: identifying the optimal individualized sequence and timing of policy interventions to enhance social welfare.

[3]As Huber (2019) highlights, the identification of instantaneous and dynamic treatment effects in our setting bears resemblance to the identification of direct and indirect causal effects in mediation analysis. In particular, the adaptive targeting effect in our context is akin to an indirect effect through multiple mediators. Nonparametric identification is known to be unattainable with only a sequential unconfoundedness assumption (Avin, Shpitser, and Pearl, 2005). We propose an identification approach for the adaptive targeting effect that utilizes an additional rank-invariance assumption (Chernozhukov and Hansen, 2005). This result may be of independent interest for the identification of the indirect effect in mediation analysis.

bounds for the welfare gains of a dynamic targeting policy. To our knowledge, the existing works of dynamic treatment regimes commonly assume away the external validity concern and have not performed robustness analysis against it. Our robustness analysis therefore adds a new conceptual and methodological contribution to the literature.

Finally, our study is also related to the multi-armed bandit literature, but the research question and policy environment differ fundamentally between our framework and the multi-armed bandit approach.[4] The multi-armed bandit method considers different individuals arriving in each period, each of whom receives treatment only once. In contrast, our study examines a policy environment in which the same set of individuals persists across periods and may receive treatments multiple times. Such policy environments are common in real-world economic policies, including job training, unemployment insurance, healthcare, education, environmental, and energy programs[5]. Our framework enables researchers and policymakers to estimate the effects of sequential treatments *ex ante*—before the allocation task—providing guidance for policy design in these policies. Hence, our framework and the multi-armed bandit approach are complementary and serve distinct objectives.

## 2 Conceptual Framework

This section introduces a conceptual framework for dynamic policy targeting. Section 2.1 formulates the dynamic treatment framework and defines dynamic targeting. Section 2.2 introduces static targeting as a benchmark policy. Section 2.3 conceptually discusses the mechanisms through which dynamic targeting can outperform static targeting.

[4]For an overview of the multi-armed bandit literature, see Lattimore and Szepesvári (2020). For recent developments of bandit algorithms relevant to economics, see Dimakopoulou, Zhou, Athey, and Imbens (2017), Ariu, Kato, Komiyama, McAlinn, and Qin (2021), Kasy and Sautmann (2021), and Kock, Preinerstorfer, and Veliyev (2022).

[5]As discussed earlier, examples include job training programs (Lechner, 2009; Rodríguez, Saltiel, and Urzúa, 2022), unemployment insurance programs (Meyer, 1995; Kolsrud, Landais, Nilsson, and Spinnewijn, 2018), healthcare programs (Luckett, Laber, Kahkoska, Maahs, Mayer-Davis, and Kosorok, 2019), and educational interventions (Ding and Lehrer, 2010)

## 2.1 Dynamic Targeting

Consider a planner who implements a policy intervention on a heterogeneous population. The planner assigns a binary treatment to individuals, taking advantage of their heterogeneous responses. Unlike the standard static treatment assignment framework studied in Manski (2004), Hirano and Porter (2009), Kitagawa and Tetenov (2018), among others, the planner can assign different treatments to the same individual across multiple periods. We focus on a two-period setting in which the planner makes treatment decisions in two time periods, $t = 1$ and $t = 2$.[6]

### 2.1.1 Dynamic Potential Outcomes

In each period, an individual is either treated ($T$) or untreated ($U$). We denote a sequence of treatment assignments over the two periods by $(d_1, d_2) \in \{U, T\}^2$. There are four possible sequential interventions: $(d_1, d_2) = (U, U)$, $(T, U)$, $(U, T)$, and $(T, T)$. For example, $(T, T)$ represents treatment in both periods, while $(U, T)$ represents treatment only in the second period.

The planner aims to maximize a social welfare criterion by assigning each individual to one of four possible sequential intervention paths. At the end of each period, the planner observes the corresponding outcomes. Let $Y_1(T)$ and $Y_1(U)$ denote the potential outcomes in period 1 when an individual is treated ($d_1 = T$) and untreated ($d_1 = U$), respectively. We assume that the potential outcomes in period 1 do not depend on the subsequent intervention $d_2$. This assumption is analogous to the no-anticipation assumption in Abbring and Van den Berg (2003). Let $Y_2(d_1, d_2)$ denote the period-2 potential outcome under the sequential intervention $(d_1, d_2) \in \{U, T\}^2$, where dependence on $d_1$ captures the dynamic effect of the initial intervention on later outcomes.

### 2.1.2 Information Available to the Planner

We define the sequential information available to the planner and the corresponding treatment assignment decisions. At the beginning of period 1, the planner observes covariates $S_1$ and assigns treatment $d_1 \in \{U, T\}$. Examples of $S_1$ include household demographics, housing characteristics,

[6] Our analysis can be readily extended to more than two periods.

and historical electricity consumption. At the end of period 1, the planner observes the response to treatment $Y_1(d_1)$.

At the beginning of period 2, the planner observes updated state information $S_2$, which depends on the first-period treatment and is thus denoted by $S_2(d_1)$. The set $S_2(d_1)$ includes state variables observed after the first-period intervention, such as the realized first-period outcome $Y_1(d_1)$ and pre-treatment electricity usage in period 2. Based on the information $(S_1, d_1, S_2(d_1))$, the planner assigns the second-period treatment.

Figure 1 shows the information flow and the sequence of treatment decisions. Let $H_t$ denote the information available to the planner prior to period $t$. At $t = 1$, the planner observes $H_1 = S_1$, and at $t = 2$, the information set expands to $H_2(d_1) = (S_1, d_1, S_2(d_1))$. To accommodate counterfactual realizations that depend on the first-period treatment, we express the period-2 information set in potential-outcome notation as $H_2(d_1)$.

The framework allows for heterogeneity in treatment responses across periods and individuals. In the presence of such heterogeneity, a uniform (non-individualized) assignment of treatment or no-treatment is generally suboptimal. Accordingly, we consider individualized sequential treatment assignments. In period 1, the planner assigns $d_1 \in \{T, U\}$ based on $H_1$; in period 2, after observing $H_2(d_1)$, the planner assigns $d_2 \in \{T, U\}$ based on $H_2(d_1)$.

We define a targeting policy in each period $t$ as a rule that the planner follows to assign treatment in period $t$ based on the available information $H_t$. Formally, a targeting policy in period $t$ is a measurable mapping $\pi_t : \mathcal{H}_t \to \{T, U\}$, where $\mathcal{H}_t$ denotes the history space in period $t$. The pair $\pi := (\pi_1, \pi_2)$ defines a dynamic targeting policy, which specifies the sequence of individualized treatment assignment rules over time.

### 2.1.3 Optimal Dynamic Targeting

We define the planner's social welfare criterion $W(\pi)$ as the expected sum of individual outcomes across the two periods, evaluated under the sequential treatment assignments specified by

$\pi$:

$$W(\pi) \equiv E\left[\sum_{(d_1,d_2)\in\{T,U\}^2} (Y_1(d_1) + Y_2(d_1,d_2)) \cdot \mathbf{1}\{\pi_1(H_1) = d_1, \pi_2(H_2(d_1)) = d_2\}\right]. \quad (1)$$

This welfare criterion aggregates flow outcomes over time, in contrast to the dynamic treatment regime literature in medical applications, which typically defines welfare only in terms of the terminal-period outcome (Robins, 1986; Murphy, 2003; Chakraborty and Moodie, 2013; Zhang, Laber, Davidian, and Tsiatis, 2018; Tsiatis, Davidian, Holloway, and Laber, 2019).

The optimal dynamic targeting policy, denoted by $\pi^* = (\pi_1^*, \pi_2^*)$, is the sequence of assignment policies that maximizes $W(\pi)$:

$$\pi^* = \underset{\pi=(\pi_1,\pi_2)\in\widetilde{\Pi}}{\arg\max}\; W(\pi), \quad (2)$$

where $\widetilde{\Pi} = \widetilde{\Pi}_1 \times \widetilde{\Pi}_2$, and $\widetilde{\Pi}_t$ denotes a set of measurable functions $\pi_t$ for period $t$.

The optimal dynamic targeting policy features two key dynamic elements. First, it allows switching between interventions $U$ and $T$ across periods—for example, assigning $(T, U)$ or $(U, T)$ instead of being restricted to $(T, T)$ or $(U, U)$—which can yield higher welfare than policies that fix a single intervention. Second, it updates treatment assignments using new information $H_2(d_1)$ that is observed after the first period. Using this sequentially revealed information, the optimal dynamic targeting can improve welfare compared to static policies based solely on initial covariates $H_1$.

## 2.2 Static Targeting

As a benchmark, we consider a static targeting policy under which the planner optimizes welfare in each period separately, without taking dynamic considerations into account. Consider that the planner has access to pre-intervention data, denoted by $H_1$, which can be used to optimize targeting in periods 1 and 2 independently. Examples of $H_1$ include baseline household characteristics and household-level historical energy consumption.

The optimal static targeting policy for the first period, denoted by $\pi_{S,1}^* : \mathcal{H}_1 \mapsto \{U, T\}$, maximizes the first-period welfare depending on $H_1$:

$$\pi_{S,1}^* = \underset{\text{all measurable } \pi_1:\mathcal{H}_1 \mapsto \{U,T\}}{\arg\max} E[Y_1(\pi_1(H_1))].$$

This policy optimizes first-period welfare without regard to its implications for second-period outcomes.

The optimal static targeting policy for the second period, denoted by $\pi_{S,2}^* : \mathcal{H}_1 \mapsto \{U, T\}$, maximizes second-period welfare under the scenario in which no intervention is applied in the first period:

$$\pi_{S,2}^* = \underset{\text{all measurable } \pi_2:\mathcal{H}_1 \mapsto \{U,T\}}{\arg\max} E[Y_2(U, \pi_2(H_1))].$$

This policy is determined independently of the first-period static policy $\pi_{S,1}^*$.

The social welfare under the static targeting policy $\pi_S^* = (\pi_{S,1}^*, \pi_{S,2}^*)$ is defined as

$$W_S(\pi_S^*) \equiv E\left[\sum_{(d_1,d_2)\in\{T,U\}^2} (Y_1(d_1) + Y_2(d_1, d_2)) \cdot \mathbf{1}\{\pi_{S,1}^*(H_1) = d_1, \pi_{S,2}^*(H_1) = d_2\}\right]. \quad (3)$$

This expression defines the social welfare realized when the sequential treatment allocation follows the static targeting policy $\pi_S$. As such, it is directly comparable to the welfare criterion used for the dynamic targeting policy.

## 2.3 Welfare Gains from Dynamic Targeting

In this section, we highlight three behavioral channels through which dynamic targeting can improve welfare—habit formation effects, learning effects, and adaptive targeting effects (i.e., the exploration of predictive information about future treatment responses). We also show that the welfare gains from dynamic targeting can be decomposed into these three channels, as formalized in Proposition 2.1.

### 2.3.1 Habit Formation Effect

Consider two other potential outcomes for the second period: $Y_2(U, U)$ and $Y_2(T, U)$. In both cases, the individual is not treated in the second period; the difference lies in the first-period treatment status. These two potential outcomes would be identical (i.e., $Y_2(T, U) - Y_2(U, U) = 0$) if the second-period untreated outcome, $Y_2(\cdot, U)$, were causally independent of the first-period treatment. However, this equality may not hold if the first-period treatment affects the second-period untreated potential outcome.

We refer to this as the *habit formation effect*, characterized by $Y_2(T, U) - Y_2(U, U) > 0$. An individual who is treated in the first period may develop habits that persist into the second period, even in the absence of treatment (Becker and Murphy, 1988).

### 2.3.2 Learning and Habituation Effects

Similarly, consider two potential outcomes for the *second* period: $Y_2(U, T)$ and $Y_2(T, T)$. In both cases, an individual receives treatment in the second period; the difference lies in the first-period treatment status. These two potential outcomes would be identical (i.e., $Y_2(T, T) - Y_2(U, T) = 0$) if the second-period treated outcome, $Y_2(\cdot, T)$, were causally independent of the first-period treatment status. However, this equality may not hold when either of the following two mechanisms is present.

The first possibility is the *learning effect*, where $Y_2(T, T) - Y_2(U, T) > 0$. An individual who learns how to respond to the intervention during the first period may exhibit a greater welfare gain in the second period than if they had not been treated initially.

The second possibility is the *habituation effect*, where $Y_2(T, T) - Y_2(U, T) < 0$. An individual treated in the first period may become habituated to the same intervention, leading to a diminished treatment response in the second period relative to having been untreated in the first period. Habituation effects are found in many empirical studies including Ito, Ida, and Tanaka (2018).

### 2.3.3 Adaptive Targeting Effect

A key advantage of dynamic targeting is its ability to optimize the second-period treatment assignment $\pi_2$ adaptively to $S_2(d_1)$—the post-first-period state information that depends on the first-period treatment $d_1$. There are two potential states of $S_2(d_1)$. If the planner assigns $d_1 = T$ in period 1, then $S_2(T)$ is observed; if $d_1 = U$, then $S_2(U)$ is observed. Consequently, the information set available to the planner differs depending on the first-period intervention, which in turn influences the prediction of households' treatment responses in period 2. For example, if $S_2(T)$ is more informative than $S_2(U)$ for predicting heterogeneity in second-period outcomes, assigning $d_1 = T$ to obtain $S_2(T)$—even at the cost of lower welfare in period 1— and letting $\pi_2$ adaptively exploit it may ultimately improve the total welfare.

We refer to this channel of welfare gain as the *adaptive targeting effect*. Given an optimal dynamic targeting $\pi^*$, the adaptive targeting effect is defined as,

$$Y_2\Big(T, \pi_2^*\big(H_2(T)\big)\Big) - Y_2\Big(T, \pi_2^*\big(H_2(U)\big)\Big), \tag{4}$$

where $H_2(T) \equiv \big(S_1, S_2(T)\big)$ denotes the second-period information set when the first-period treatment is $d_1 = T$, and $H_2(U) \equiv \big(S_1, S_2(U)\big)$ denotes the corresponding information set when $d_1 = U$.

Given an optimal dynamic targeting policy $\pi^*$, the source of the adaptive targeting effect is through the acquisition of households' responses to the previous treatment. A positive adaptive targeting effect indicates that $H_2(T)$ provides more informative signals for determining the optimal second-period intervention $\pi_2^*$. In such cases, assigning $d_1 = T$ in the first period may be justified, even if it entails lower welfare in period 1. The adaptive targeting effect is zero when the optimal second-period assignment remains unchanged across the two information sets, that is, when $\pi_2^*\big(H_2(T)\big) = \pi_2^*\big(H_2(U)\big)$. Finally, the adaptive targeting effect is negative if $H_2(T)$ provides *less* informative signals than $H_2(U)$ for determining $\pi_2^*$. In Section 5.3, we empirically estimate the adaptive targeting effects in our experiment, showing that we indeed find positive, zero, and negative adaptive targeting effects, depending on households.

#### 2.3.4 Decomposition of Welfare Gains

Proposition 2.1 shows that the welfare gain from dynamic targeting can be decomposed into five components, with the proof provided in Appendix A.2.

**Proposition 2.1.** *The welfare gain from the optimal dynamic targeting policy relative to no intervention, $W(\pi^*) - W_{UU}$, where $W_{UU} \equiv E[Y_1(U) + Y_2(U,U)]$, can be decomposed as follows:*

$$
\begin{aligned}
& W(\pi^*) - W_{UU} \\
&= \underbrace{E[Y_1(T) - Y_1(U) | \pi_1^*(H_1) = T]}_{\text{Treatment effect on the treated in } t=1} \cdot \Pr(\pi_1^*(H_1) = T) \\
&+ \underbrace{E[Y_2(U,T) - Y_2(U,U) | \pi_2^*(H_2(U)) = T]}_{\text{Treatment effect on the treated in } t=2} \cdot \Pr(\pi_2^*(H_2(U)) = T) \\
&+ \underbrace{E[Y_2(T,U) - Y_2(U,U) | \pi_1^*(H_1) = T, \pi_2^*(H_2(T)) = U]}_{\text{Habit formation effect for those assigned to } (T,U)} \cdot \Pr(\pi_1^*(H_1) = T, \pi_2^*(H_2(T)) = U) \\
&+ \underbrace{E[Y_2(T,T) - Y_2(U,T) | \pi_1^*(H_1) = T, \pi_2^*(H_2(T)) = T]}_{\text{Learning effect for those assigned to } (T,T)} \cdot \Pr(\pi_1^*(H_1) = T, \pi_2^*(H_2(T)) = T) \\
&+ \underbrace{E[Y_2(T, \pi_2^*(H_2(T))) - Y_2(T, \pi_2^*(H_2(U))) | \pi_1^*(H_1) = T]}_{\text{Adaptive Targeting Effect on the treated in } t=1} \cdot \Pr(\pi_1^*(H_1) = T). \qquad (5)
\end{aligned}
$$

This decomposition shows that the welfare gain from the optimal dynamic targeting policy, $\pi^*$, can be expressed as the sum of five distinct components: (i) the average first-period treatment effect among individuals treated in period 1, (ii) the average second-period treatment effect among those treated in period 2, (iii) the average habit-formation effect among those assigned to $(T,U)$, (iv) the average learning/habituation effect among those assigned to $(T,T)$, and (v) the average adaptive targeting effect among those treated in the first period. Each component is defined over the intervention groups induced by the optimal policy $\pi^*$ and weighted by the relative frequencies of the corresponding conditioning groups.

The proof of this decomposition does not rely on the optimality of $\pi$; that is, it applies not only to $\pi^*$ but to any dynamic targeting policy $\pi$ (see Appendix A.2). In Section 5, we develop an econometric framework to estimate each component of the decomposition and empirically evaluate them using our experimental data.

Finally, a similar decomposition can be established for the optimal static targeting policy $\pi_S^*$ (see Appendix A.3). This decomposition consists of the same four effects but excludes the adaptive targeting effect, as static targeting does not utilize the updated information $H_2(d_1)$ when determining the second-period treatment. This highlights that only dynamic targeting can exploit the adaptive targeting effect to improve welfare.

# 3 Field Experiment and Data

This section describes the design and implementation of our two-period sequential RCT in the context of a residential energy rebate program in Japan. Section 3.1 outlines the field experiments. Section 3.2 presents the summary statistics and balance test results.

## 3.1 Field Experiment

In the summer and winter of 2020, we conducted a two-period sequential RCT in the field in the Kansai and Chubu regions of Japan, in collaboration with the Japanese Ministry of the Environment.[7] We recruited a diverse group of 2,400 households via letter and e-mail, offering a participation fee of 2,000 JPY (≈20 USD; 1¢≈1 JPY in 2020).[8]

The experiment followed a two-period sequential RCT. We randomly assigned our customers to one of four groups: $(U, U)$, untreated in both the first period (summer 2020) and the second

[7]The experiment received approval from the ethics committee of the Inter-Graduate School Program for Sustainable Development and Survivable Societies at Kyoto University and was registered in the AEA RCT Registry (Ida, Ishihara, Kido, and Sasaki, 2020).

[8]Because our experiment was an RCT for households who agreed to participate in the experiment, the external validity of the sample is an important question. To investigate this point, we collected data from a random sample of 2,070 customers who resided in the experimental locations and show the summary statistics in Table A.1. We find that the observable characteristics of the two groups are not largely different, but the experimental sample has slightly higher sample averages in their monthly electricity usage, number of people at home on weekdays, self-efficacy in energy conservation, and household income.

period (winter 2020); $(U, T)$, untreated in the first period and treated in the second; $(T, U)$, treated in the first period and untreated in the second; and $(T, T)$, treated in both periods. The number of households in each group is 625 for $(U, U)$, 606 for $(U, T)$, 581 for $(T, U)$, and 588 for $(T, T)$.

The treatment of our experiment was a "peak-time rebate" (PTR) program for residential electricity conservation (Wolak, 2011; Ida, Ishihara, Ito, Kido, Kitagawa, Sakaguchi, and Sasaki, 2026). The fundamental inefficiency in many electricity markets is the failure of residential electricity prices to reflect time-varying marginal costs. During peak hours, the invariant residential price is often too low compared to the variant marginal cost, leading to a short-run deadweight loss. PTR programs aim to reduce this loss by aligning the rebate incentive with the marginal cost.

The goal of our PTR was to reduce residential electricity consumption during system peak hours—1:00–5:00 p.m. in summer and 5:00-9:00 p.m. in winter—over two treatment weeks: August 24–30, 2020 (period 1) and December 14–20, 2020 (period 2). During these treatment weeks, customers participating in the rebate program received payments equal to their electricity savings during peak hours relative to their baseline usage (in kWh), multiplied by 100 JPY per kWh. This rebate rate was approximately equal to the wholesale electricity market price during peak hours. Baseline consumption was calculated as each customer's average peak-hour usage before the interventions, ensuring that customers could not deliberately alter their baseline.[9]

## 3.2 Data and Summary Statistics

We collected household-level electricity consumption data in 30-minute intervals during the pre-experimental period (July 1–31 and November 11–30, 2020) and the experimental period (August 24–30 and December 14–20, 2020). Before the experiment, a survey was conducted to gather various household characteristics.

Table 1 provides summary statistics and a balance check.[10] Columns 1–4 display the sample

[9] Before the interventions, we recorded each customer's hourly electricity usage and calculated their average hourly consumption during the relevant peak hours—1 pm to 5 pm for the summer interventions and 5 pm to 9 pm for the winter interventions. We did not inform customers how their baseline was calculated until treatment assignments were announced, ensuring that they had no opportunity to deliberately alter their baseline consumption.

[10] Appendix A.6 presents empirical evidence of the average treatment effects and the dynamic heterogeneity in the

averages for each random assignment group in each period, with standard deviations in brackets.

The initial six variables represent pre-experimental electricity usage (in watt-hours per 30-minute interval) during peak hours (summer: 1 PM to 5 PM; winter: 5 PM to 9 PM), pre-peak hours (summer: 10 AM to 1 PM; winter: 2 PM to 5 PM), and post-peak hours (summer: 5 PM to 8 PM; winter: 9 PM to midnight). The remaining variables, derived from the survey, include the number of household members typically at home on weekdays and self-efficacy in energy conservation, measured on a 5-point Likert scale (with higher scores indicating greater self-efficacy). Household income was reported in units of 1000 JPY. Table 1 verifies the balance of all variables across the randomly-assigned groups.

# 4 Optimal Assignment Policy and Welfare Gains

This section empirically evaluates the optimal dynamic targeting policy using sequential experimental data from an energy-saving rebate program. Section 4.1 describes the construction of the social welfare criterion, Section 4.2 outlines the estimation of static and dynamic targeting rules, Section 4.3 presents the results, and Section 4.4 examines potential dynamic manipulation and its welfare implications.

## 4.1 Social Welfare Criterion

Let $p$ denote the retail electricity price and $c$ the marginal cost of electricity. During peak hours, the time-invariant residential price $p$ is typically below the marginal cost $c$. The goal of rebate programs is to reduce the resulting welfare loss by aligning marginal rebate incentives with $c$. For each sequential intervention $(d_1, d_2) \in \{T, U\}^2$, let $Q_1(d_1)$ and $Q_2(d_1, d_2)$ denote the potential outcomes of electricity consumption in periods 1 and 2. Assuming a locally linear demand curve, the short-run welfare contributions of intervention $(d_1, d_2)$ can be expressed as $\frac{1}{2}(p-c)\big(Q_1(d_1)-Q_1(U)\big)$ and $\frac{1}{2}(p-c)\big(Q_2(d_1,d_2)-Q_2(U,U)\big)$. We also incorporate the long-run

treatment effects on peak-hour electricity usage, suggesting that dynamic policy targeting, as discussed in Section 4, could significantly enhance the policy outcome.

welfare gain from reduced consumption using the capacity market price per kW ($\delta$), which reflects avoided investment in generation capacity. Finally, each period of program participation involves an implementation cost $a$ per customer.

The social welfare gain from each household in periods 1 and 2 ($Y_1$ and $Y_2$) can be expressed as $Y_1(d_1) \equiv b \cdot (Q_1(d_1) - Q_1(U)) - a \cdot \mathbf{1}\{d_1 = T\}$ and $Y_2(d_1, d_2) \equiv b \cdot (Q_2(d_1, d_2) - Q_2(U, U)) - a \cdot \mathbf{1}\{d_2 = T\}$, where $b = \frac{1}{2}(p - c) + \delta$. Note that no treatment in period 1 yields zero welfare contribution in that period (i.e., $Y_1(U) = 0$), and no treatment in both periods yields zero welfare contribution in period 2 (i.e., $Y_2(U, U) = 0$). We then substitute $Y_1(d_1)$ and $Y_2(d_1, d_2)$ into equations (1) and (3) to construct the social welfare functions under static targeting $W(\pi_S)$ and dynamic targeting $W(\pi)$.[11]

## 4.2 Estimation of Static and Dynamic Targeting Policies

To estimate the optimal dynamic and static targeting policies defined in Section 2, we build on the Empirical Welfare Maximization (EWM) method of Kitagawa and Tetenov (2018) and its dynamic extension proposed by Sakaguchi (2025). From the field experiment described in Section 3.1, we have sequential RCT data: $\{D_{it}, Y_{it}, S_{it} : i = 1, \ldots, n; t = 1, 2\}$, where $D_{i1}$ and $D_{i2}$ are randomly assigned binary treatments for household $i$ in periods 1 and 2, and the outcome variable $Y_{it}$ is the observed welfare contribution defined in Section 4.1.

*Dynamic targeting policy:* Let $\Pi \equiv \Pi_1 \times \Pi_2$ denote a pre-specified class of dynamic targeting policies, where $\Pi_1$ and $\Pi_2$ represent the sets of feasible policies in periods 1 and 2. We use a backward induction method, first estimating the optimal second-period policy $\pi_2^*$ by solving the EWM problem for period 2:

$$\hat{\pi}_2^* \in \arg\max_{\pi_2 \in \Pi_2} \widehat{W}_2(\pi_2), \quad \widehat{W}_2(\pi_2) \equiv \frac{1}{n} \sum_{i=1}^{n} \frac{Y_{i2} \cdot \mathbf{1}\{D_{i2} = \pi_2(H_{i2})\}}{P(D_2 = D_{i2} | H_2 = H_{i2})},$$

where $\widehat{W}_2(\pi_2)$ is the empirical welfare function in period 2. The propensity score $P(D_2 = D_{i2} \mid$

[11] In Appendix A.5, we provide a description of the exogenous parameters in the social welfare function. We use the same values of $p, c, \delta$ for both periods, but this assumption is not necessary for our framework.

$H_2 = H_{i2})$ can be obtained from the RCT data.

Given $\hat{\pi}_2^*$, we estimate the optimal first-period policy $\pi_1^*$ by solving the EWM problem:

$$\hat{\pi}_1^* \in \arg\max_{\pi_1 \in \Pi_1} \widehat{W}(\pi_1), \quad \widehat{W}(\pi_1) \equiv \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbf{1}\{D_1 = \pi_1(H_{i1})\}}{P(D_1 = D_{i1} | H_1 = H_{i1})} \left( Y_{i1} + \frac{Y_{i2} \cdot \mathbf{1}\{D_{i2} = \hat{\pi}_2^*(H_{i2})\}}{P(D_2 = D_{i2} | H_2 = H_{i2})} \right),$$

where $\widehat{W}(\pi_1)$ is the empirical welfare function of $\pi_1$ given the second-period policy $\hat{\pi}_2^*$. Using an iterative procedure, we obtain the estimated dynamic targeting policy $\hat{\pi}^* = (\hat{\pi}_1^*, \hat{\pi}_2^*)$. Sakaguchi (2025) shows that the resulting estimator $\hat{\pi}^*$ is consistent and rate-optimal for the true optimal dynamic targeting policy $\pi^*$, provided that the policy class $\Pi$ includes the optimal $\pi^*$.

An advantage of the EWM method is that it is model-free, requiring no parametric or functional form assumptions on the potential outcome distributions. The policy class $\Pi_t$ for each $t$, however, must be specified. If the policy class is overly flexible, the estimated policy $\hat{\pi}^*$ may overfit the data, yielding lower welfare when implemented. To address this concern, we use decision tree classes (Breiman, Friedman, Olshen, and Stone, 2017), which mitigate overfitting by controlling tree depth and leverage efficient search algorithms developed in the classification tree literature.

Specifically, we employ the class of depth-4 decision trees for each of $\Pi_1$ and $\Pi_2$. In each step of the dynamic EWM, we maximize the empirical welfare criterion within this policy class using the exhaustive search algorithm of Zhou, Athey, and Wager (2023).[12]

*Static targeting policy:* Let $\Pi_S = \Pi_{S,1} \times \Pi_{S,2}$ denote a pre-specified class of static targeting policies. As in the estimation of the optimal dynamic targeting policy, we employ the class of depth-4 decision trees for each $\Pi_{S,t}$. Since static targeting is a myopically optimizing policy, its estimation differs from that of dynamic targeting. For period 1, the first-period targeting $\pi_{S,1}^*$ is

[12]Because obtaining a globally optimal depth-4 tree is computationally demanding, we use a heuristic two-step approximation. We first estimate a depth-2 "parent" tree that maximizes empirical welfare in the full sample, dividing it into four subsamples. For each subsample, we then estimate a depth-2 "child" tree that maximizes welfare within that subset and graft the resulting four child trees onto the parent. This grafted-tree approach is a standard technique in the machine learning literature for constructing deep trees under computational constraints (see, e.g., Chapter 2 of Breiman, Friedman, Olshen, and Stone (2017) and Section 9.2 of Hastie, Tibshirani, and Friedman (2009)).

estimated by EWM as

$$\hat{\pi}^*_{S,1} \in \arg\max_{\pi_1 \in \Pi_1} \widehat{W}_1(\pi_1), \quad \widehat{W}_1(\pi_1) \equiv \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_{i1} \cdot \mathbf{1}\{\pi_1(H_{i1}) = D_{i1}\}}{P(D_1 = D_{i1}|H_1 = H_{i1})}\right),$$

where $\widehat{W}_1(\pi_1)$ is the empirical welfare function of $\pi_1$ for period 1. Separately, for period 2, the second-period targeting is estimated as

$$\hat{\pi}^*_{S,2} \in \arg\max_{\pi_2 \in \Pi_2} \widehat{W}_2(U, \pi_2), \quad \widehat{W}_2(U, \pi_2) \equiv \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_{i2} \cdot \mathbf{1}\{D_{i1} = U, D_{i2} = \pi_2(H_{i1})\}}{P(D_1 = D_{i1}|H_1 = H_{i1})P(D_2 = D_{i2}|H_2 = H_{i2})}\right),$$

where $\widehat{W}_2(U, \pi_2)$ is the empirical welfare function of $\pi_2$ for period 2 under the counterfactual scenario that no one is treated in the period 1.

Once the optimal policies are estimated, we estimate the welfare gains under each policy assignment: $\Delta W(\pi^*) \equiv W(\pi^*) - W_{UU}$ and $\Delta W_S\big(\pi^*_S\big) \equiv W_S(\pi^*_s) - W_{UU}$. A known caveat of EWM estimation is that the optimized empirical welfare values, $\widehat{W}(\hat{\pi}^*)$ and $\widehat{W}_S(\hat{\pi}^*_S)$, tend to be an upwardly biased estimate of the true optimal welfare. This bias, often referred to as the winner's curse or winner's bias, arises because the same data are used to both estimate the policy and evaluate its welfare (Andrews, Kitagawa, and McCloskey, 2019). To mitigate this bias, we follow Ida, Ishihara, Ito, Kido, Kitagawa, Sakaguchi, and Sasaki (2026) and generate pseudo-samples by permuting regression residuals obtained from a causal forest fitted on all covariates, thereby constructing an approximately unbiased test dataset for welfare evaluation (see Appendix A.7 for details).

## 4.3 Estimation Results

We apply the estimation methods described in Section 4.2 to our sequential RCT data: $\{D_{it}, Y_{it}, S_{it} : i = 1, \ldots, n; t = 1, 2\}$, where $D_{i1}$ and $D_{i2}$ are randomly assigned binary treatments for household $i$ in periods 1 and 2, and the outcome variable $Y_{it}$ is the observed welfare contribution defined in Section 4.1.

We use five pre-intervention state variables $S_{i1}$: (i) peak-time baseline electricity consumption before the experiment, (ii) household income, (iii) number of household members at home from 13:00–17:00 on weekdays, (iv) number of members at home from 17:00–21:00 on weekdays, and (v) household self-efficacy in energy conservation. Likewise, we use two variables as the post–first-period state variables $S_{i2}$: (i) peak-time electricity consumption during the first-period event ($Q_{i1}$) and (ii) peak-time baseline electricity consumption in winter. These variables are selected based on their predictive power for electricity consumption and treatment effect heterogeneity.[13]

In Table 2 and Figure 2, we begin by presenting the welfare gains from four non-targeting policies: 100% $(U, U)$, 100% $(T, U)$, 100% $(U, T)$, and 100% $(T, T)$. These policies assign all households to the same treatment sequence without targeting. For example, the 100% $(U, T)$ policy treats no households in period 1 and all households in period 2. The point estimates indicate that non-targeting policies involving at least one treatment generate positive welfare gains; however, these gains are not statistically different from zero at conventional significance levels.

We then report results for the optimal static targeting policy, $\pi_S^*$. As defined in Section 2.2, the static targeting policy myopically maximizes welfare in each period. Our results show that even the myopic static targeting policy, $\pi_S^*$, yields substantially larger welfare gains than any of the non-targeting policies ($856.7$ JPY), indicating that targeting improves welfare even when it is implemented myopically.

Finally, the dynamic targeting policy $\pi^*$ incorporates both pre-experimental information ($H_1$) and post–first-period information ($H_2$) to adaptively assign each household to one of the four possible treatment sequences. Although $\pi_S^*$ and $\pi^*$ share the same flexibility in the set of feasible treatment sequences, the dynamic targeting policy $\pi^*$ delivers substantially larger welfare gains ($1665.4$ JPY). This finding highlights that adaptively assigning treatments based on post–first-period information ($H_2$) markedly enhances social welfare.

Table 2 also reports the share of households assigned to each of the four possible treatment

[13] To improve predictive power, we specify the second-period policy class $\Pi_2$ so that each policy $\pi_2 \in \Pi_2$ bases the second-period treatment decision on the difference between peak and baseline electricity usage in period 1 (i.e., the energy savings), rather than on baseline electricity usage itself.

sequences. Under the static targeting policy $\pi_S^*$, the majority of households are treated in period 1. This pattern arises because myopic maximization of first-period welfare assigns treatment to a large fraction of households in the first period. In contrast, the dynamic targeting policy assigns households more diversely across the four treatment sequences, with fewer households treated in period 1. This shift accounts for the additional welfare gains relative to $\pi_S^*$, where the dynamic targeting policy $\pi^*$ further reshapes the distribution of treatment sequences by adaptively assigning treatments. In Section 2.3, we examine this point in more detail and investigate the mechanisms through which $\pi^*$ generates a large welfare gain.

Figure 3 further compares the treatment assignment ratios across the four sequential interventions under the optimal dynamic and static targeting policies. The results reveal substantial differences in treatment assignments between the two approaches, showing that dynamic targeting reallocates many households across intervention sequences relative to static targeting. This pattern indicates that dynamic and static targeting rely on distinct assignment mechanisms to enhance welfare.

In Table 3, we compare welfare gains across alternative policies and further decompose the welfare differences into periods 1 and 2. The results show that the dynamic targeting policy yields significantly higher total welfare gains than both the non-targeting and static targeting policies.

Focusing on period-specific welfare differences, the dynamic targeting policy $\pi^*$ generates lower welfare gains than the static targeting policy $\pi_S^*$ in period 1 ($-280.9$ JPY), but higher welfare gains in period 2 ($1089.6$ JPY), resulting in a larger total welfare gain overall ($808.7$ JPY). This pattern indicates that the primary advantage of dynamic targeting arises from adaptive treatment assignment in period 2, which exploits new information revealed by the period-1 intervention, even though it deliberately sacrifices some welfare gains in period 1.

## 4.4 Potential Anticipatory Responses and Welfare Implications

Households in our experiment were informed that their treatments were randomly assigned independent of their past consumption. As a result, they had no incentive to manipulate their first-period outcomes so as to receive their preferred treatments in the second period. However, when the planner implements the estimated policy, customers subject to the targeting rule may have an incentive to manipulate their first-period consumption to influence their second-period treatment assignment if they know the dynamic targeting rule.

In this section, we assess how potential manipulation of first-period consumption arising from such anticipatory responses affects the robustness of the welfare gains from dynamic targeting. Specifically, we construct bounds on the welfare gain under the estimated policy in extreme scenarios in which all households perfectly know the dynamic targeting rule and manipulate their first-period consumption to secure second-period treatment.[14] This behavior affects total welfare through two channels: first, by altering electricity consumption in the first period, and second, by changing treatment status in the second period and the resulting second-period consumption.

For each household $i$ in the sample, let $\tilde{Q}_i(d_1)$ denote period-1 consumption under treatment $d_1$ when the assignment rule $\hat{\pi}$ is known. Due to anticipatory responses, $\tilde{Q}_i(d_1)$ may differ from $Q_1(d_1)$, which represents consumption under sequentially randomized treatment.

Because the RCT sample does not provide direct observations of $\tilde{Q}_i(d_1)$, we construct $\tilde{Q}_i(d_1)$ for each $i$ as follows. If $\hat{\pi}_2$ assigns $i$ to treatment based on $Q_i(\hat{\pi}_1(H_1))$, $i$ does not have an incentive to manipulate even with the knowledge of $\hat{\pi}_2$ so that we set $\tilde{Q}_i(\hat{\pi}_1(H_1)) = Q_i(\hat{\pi}_1(H_1))$. In contrast, if $\hat{\pi}_2$ assigns $i$ to non-treatment based on $Q_i(\hat{\pi}_1(H_1))$, $i$ has an incentive to manipulate the period-1 consumption to receive treatment in the second period. In this case, we assume that $\tilde{Q}_i(\hat{\pi}_1(H_1))$ differs from $Q_i(\hat{\pi}_1(H_1))$ by the amount minimally sufficient to be treated in period 2. In our sample, approximately 8.2% of households with $\hat{\pi}_2(H_1) = U$ can have $\tilde{Q}_i(\hat{\pi}_1(H_1)) \neq Q_i(\hat{\pi}_1(H_1))$. If $i$ manipulates, whether $\tilde{Q}_i(\hat{\pi}_1(H_1))$ is smaller or larger than $Q_i(\hat{\pi}_1(H_1))$ is determined by how $\hat{\pi}_2(\cdot)$ depends on the period-1 consumption.

Whether $\tilde{Q}_i(\hat{\pi}_1(H_1))$ is smaller or larger than $Q_i(\hat{\pi}_1(H_1))$ determines whether the manipulat-

[14] In practice, it is unlikely that typical households can engage in such manipulation because the targeting rules are not straightforward. As we explain below, some households would need to increase their first-period consumption to receive treatment in the second period, whereas others would need to do the opposite, since the targeting rule depends on multidimensional household characteristics. Nevertheless, in this section we consider an extreme scenario in which all households fully understand the dynamic targeting rule and examine its welfare implications.

ing households are classified as *downward manipulators* or *upward manipulators*, respectively.[15] By further reducing peak-hour electricity consumption in period 1, downward manipulators contribute positively to welfare gains. In contrast, upward manipulators reduce welfare because their increased peak-hour electricity usage in period 1 lowers social welfare. Among households who can manipulate, 94.3% are downward manipulators and 5.7% are upward manipulators.

To generate bounds for welfare gains, we consider the following three scenarios: (i) only the downward manipulators respond to $\hat{\pi}_2$, (ii) only the upward manipulators respond to $\hat{\pi}_2$, and (iii) both types respond to $\hat{\pi}_2$. Scenario (i) leads to the upper bound of welfare gain, while Scenario (ii) leads to the lower bound in our bound analysis. We evaluate the welfare for each case using the test sample generated for policy evaluation. For the downward or upward manipulators, since the second-period treatment status switches to the counterfactual one, the corresponding period-2 potential outcomes are not observed in the original sample; we therefore impute them using random forest predictors based on their $H_2$ excluding $\tilde{Q}_1$.

Table 4 reports the results under the three manipulation scenarios. Across all scenarios, manipulation increases the number of households receiving treatment in period 2, but the welfare gains from dynamic targeting remain close to the no-manipulation benchmark, with only modest differences. When only downward manipulators adjust, total welfare increases slightly, as they reduce peak-hour electricity usage in period 1. In contrast, upward manipulation leads to a small reduction in welfare because it increases peak-hour electricity usage in period 1. Overall, these results suggest that although potential anticipatory responses is conceptually important, it does not materially affect our results in the current context.

# 5 Mechanism Behind the Dynamic Targeting

The results in Section 4 show that dynamic targeting substantially enhances social welfare relative to both non-targeting and static targeting policies. In this section, we examine the mech-

[15] Under the estimated policy, no household can qualify through both upward and downward manipulation.

anisms underlying the welfare gains from dynamic targeting through the decomposition theorem developed in Proposition 2.1.

Proposition 2.1 decomposes the welfare gains from dynamic targeting into five components: the first-period treatment effect, the second-period treatment effect, the habit-formation effect, the learning effect, and the adaptive targeting effect. A key empirical challenge is that some elements of this decomposition—the conditional averages of the habit-formation, learning, and adaptive targeting effects in equation (5)—are *counterfactual* outcomes, and methods for identifying these components have not yet been developed in the literature.

In Section 5.1, we propose a method to identify these counterfactual components. Section 5.2 then presents the corresponding estimation results, providing insights into the mechanisms underlying the welfare gains from dynamic targeting.

## 5.1 Identification

We aim to identify and estimate each component of the decomposition in equation (5). However, the following components are counterfactual and cannot be identified without additional assumptions:

- (conditional average habit-formation effect) $E[Y_2(T,U)-Y_2(U,U)|\pi_1^*(H_1)=T,\pi_2^*(H_2(T))=U]$;
- (conditional average learning effect) $E[Y_2(T,T)-Y_2(U,T)|\pi_1^*(H_1)=T,\pi_2^*(H_2(T))=T]$;
- (conditional average adaptive targeting effect) $E[Y_2(T,\pi_2^*(H_2(T)))-Y_2(T,\pi_2^*(H_2(U)))|\pi_1^*(H_1)=T]$.

To see the identification challenge, consider the habit-formation effect. Although the unconditional average effect $E[Y_2(T,U)-Y_2(U,U)]$ is identified from the randomized groups $(T,U)$ and $(U,U)$, the corresponding *conditional* effect is not. The conditioning event depends on $H_2(T)$, which is unobserved for households assigned $U$ in period 1. As a result, the counterfactual expectation $E[Y_2(U,U) \mid \pi_1^*(H_1)=T,\pi_2^*(H_2(T))=U]$ cannot be directly recovered from the data. The same reasoning applies to the conditional learning effect and the conditional adaptive targeting effect. Because each effect involves counterfactual history, they are not identified without additional assumptions.

These non-identification results correspond to what are referred to as *impossibility results* in the mediation analysis literature (Avin, Shpitser, and Pearl, 2005; Huber, 2019). To identify these components, we propose a method that relies on the *conditional rank invariance assumption* (Chernozhukov and Hansen, 2005) for the state variables $S_2(d_1)$.

**Assumption 5.1.** *(Conditional rank invariance) Let $K$ be the dimension of $S_2$, and we denote by $S_2^k(d_1)$ the $k$-th element of $S_2(d_1)$.*[16] *For each $k$, there exists a function $\psi_k$ and random variable $\varepsilon_k$ such that $S_2^k(d_1) = \psi_k(d_1, S_1, \varepsilon_k)$ a.s. for each $d_1 \in \{U, T\}$, where $\psi_k$ is continuous and strictly increasing in $\varepsilon_k$, and the conditional cumulative distribution function $F_{\varepsilon_k|S_1}(\cdot|s_1)$ is continuous on $\mathbb{R}$.*

The rank-invariance assumption is widely used in the literature to identify distributional causal parameters, such as quantile treatment effects (Heckman, Ichimura, and Todd, 1997; Chernozhukov and Hansen, 2005; Vuong and Xu, 2017).[17] This assumption requires that, conditional on $S_1$, the relative rank (or quantile) of $S_2^k(T)$ coincides with that of $S_2^k(U)$.

Assumption 5.1 holds when the error term enters additively, as is commonly assumed in applied economics research (i.e., $S_2^k(d_1) = \psi_k^*(d_1, S_1) + \varepsilon_k$ for some real-valued function $\psi_k^*$). Moreover, conditioning on a rich set of pre-intervention covariates $S_1$ enhances the plausibility of rank invariance by capturing systematic heterogeneity that may affect households' outcomes under both treatment and control. In our empirical setting, $S_1$ includes detailed pre-treatment information such as hourly electricity usage patterns, demographic characteristics, and housing characteristics, which are key determinants of electricity demand and its responsiveness to interventions. Consequently, conditional on these pre-treatment variables, it is plausible that a household's relative rank of $S_2^k(T)$ coincides with that of $S_2^k(U)$.

[16] In our empirical context, $K = 2$.

[17] Chernozhukov and Hansen (2005) apply rank invariance to identify quantile treatment effects. Vuong and Xu (2017) use it to identify individual treatment effects.

Under Assumption 5.1, the counterfactual state variable $S_2^k(d_1)$ can be expressed as

$$S_2^k(d_1) = Q_{S_2^k(d_1)|S_1}\left(F_{S_2^k(d_1')|S_1}\left(S_2^k(d_1')|S_1\right)|S_1\right), \tag{6}$$

for any $d_1, d_1' \in \{U, T\}$, where $Q_{S_2^k(d_1)|S_1}(\cdot|S_1)$ is the conditional quantile function of $S_2^k(d_1)$ given $S_1$, and $F_{S_2^k(d_1')|S_1}(\cdot|S_1)$ is the conditional distribution function of $S_2^k(d_1')$ given $S_1$.[18] Thus, $S_2^k(T)$ and $S_2^k(U)$ are linked by a one-to-one mapping, referred to as a *counterfactual mapping* (Vuong and Xu, 2017). This mapping allows us to recover counterfactual intermediate electricity use from its observed counterpart.

Under the random assignment of $(D_1, D_2)$ from our sequential RCT, the conditional distribution and quantile functions in equation (6) are identified as follows:

$$F_{S_2^k(d_1')|S_1}(\cdot|s_1) = F_{S_2^k|(D_1,S_1)}(\cdot|d_1', s_1) \text{ and } Q_{S_2^k(d_1)|S_1}(\cdot|s_1) = Q_{S_2^k|(D_1,S_1)}(\cdot|d_1, s_1).$$

Substituting these results into equation (6) identifies the distribution of $S_2^k(d_1)$, even for individuals who did not receive $d_1$ in the first period. Building on these results, Proposition 5.1 establishes the identification of the counterfactual effects in equation (5) under the assumptions of rank invariance and random assignment of the sequential interventions $(D_1, D_2)$.

**Proposition 5.1.** *Suppose that Assumption 5.1 and the following conditions hold: For each* $(d_1, d_2) \in \{U, T\}^2$, *(i)* $(Y_2(d_1, d_2), S_2(d_1)) \perp\!\!\!\perp D_1|H_1$ *and (ii)* $Y_2(d_1, d_2) \perp\!\!\!\perp D_2|H_2$. *Define* $\widetilde{H}_2(d_1) = (S_1, d_1, \widetilde{S}_2(d_1))$, *where* $\widetilde{S}_2(d_1) = (\widetilde{S}_2^1(d_1), \ldots, \widetilde{S}_2^K(d_1))$ *with*

$$\widetilde{S}_2^k(d_1) := \mathbf{1}\{D_1 = d_1\} \cdot S_2^k + \mathbf{1}\{D_1 \neq d_1\} \cdot Q_{S_2^k|(D_1,S_1)}\left(F_{S_2^k|(D_1,S_1)}(S_2^k|D_1, s_1)|d_1, s_1\right),$$

*for* $k = 1, 2, \ldots, K$. *Let* $\omega(d_1, d_2) := \mathbf{1}\{D_1 = d_1, D_2 = d_2\}/(P(D_1 = d_1|H_1) \cdot P(D_2 = d_2|H_2))$.

[18] For random variables $A$ and $B$, the conditional quantile function is defined as $Q_{A|B}(\tau|b) = \inf\{a \in \mathbb{R} : P(A \leq a|b) \geq \tau\}$ for any $\tau \in [0, 1]$.

*Then, the following identification results hold:*

$$E[Y_2(T,U) - Y_2(U,U)|\pi_1^*(H_1) = T, \pi_2^*(H_2(T)) = U] = E\left[\omega(T,U)\cdot Y_2|\pi_1^*(H_1) = T, \pi_2^*(H_2) = U\right] - E\left[\omega(U,U)\cdot Y_2 \middle| \pi_1^*(H_1) = T, \pi_2^*\left(\widetilde{H}_2(T)\right) = U\right], \tag{7}$$

$$E[Y_2(T,T) - Y_2(U,T)|\pi_1^*(H_1) = T, \pi_2^*(H_2(T)) = T] = E\left[\omega(T,T)\cdot Y_2|\pi_1^*(H_1) = T, \pi_2^*(H_2) = T\right] - E\left[\omega(U,T)\cdot Y_2 \middle| \pi_1^*(H_1) = T, \pi_2^*\left(\widetilde{H}_2(T)\right) = T\right], \tag{8}$$

$$E[Y_2\left(T, \pi_2^*\left(H_2(T)\right)\right) - Y_2\left(T, \pi_2^*\left(H_2(U)\right)\right)|\pi_1^*(H_1) = T] = E\left[\omega(T, \pi_2^*(H_2))\cdot Y_2|\pi_1^*(H_1) = T\right] - E\left[\omega\big(T, \pi_2^*\big(\widetilde{H}_2(U)\big)\big)\cdot Y_2 \middle| \pi_1^*(H_1) = T\right]. \tag{9}$$

*Proof.* See Appendix A.8. □

Conditions (i) and (ii) in Proposition 5.1 require conditional random assignment of $D_1$ and $D_2$, which is satisfied in our experiment. The proposition establishes identification of the three counterfactual components in the decomposition (5): the conditional averages of the habit-formation effect, the learning effect, and the adaptive targeting effect. These results provide a constructive basis for estimation of these components; specifically, they can be estimated with empirical analogue of equations-(7)–(9), together with quantile and distribution estimators of $Q_{S_2^k(d_1)|S_1}$ and $F_{S_2^k(d_1)|S_1}$. Appendix A.1 provides further estimation details. The remaining components in equation (5) are directly identifiable and estimable under random assignment of $(D_1, D_2)$.

## 5.2 Estimation Results on the Mechanisms Behind Dynamic Targeting

Table 5 reports the estimation results of welfare gain decomposition (5), based on the identification method described in Proposition 5.1. The conditional effects reported in column 2 correspond to the conditional expectations (e.g., $E[Y_1(T) - Y_1(U) \mid \pi_1^*(H_1) = T]$ for the 1st-period treatment effect) in equation (5), while the fractions in column 3 represent the corresponding proportions of individuals (e.g., $\Pr(\pi_1^*(H_1) = T)$) in the same equation. Column 4 then reports the product of these two terms.

All five components contribute positively to the welfare gain of the dynamic targeting $\pi^*$, indicating that the dynamic targeting leverages all of these mechanisms to enhance welfare. As described in Section 2, a distinctive feature of dynamic targeting is that it exploits the adaptive targeting effect—treating an individual in period 1 could provide valuable information to decide this individual's optimal treatment assignment in period 2. Our result indicates that the adaptive targeting effect makes a sizable positive contribution ($367.0$ JPY), suggesting that drawing on the information $H_2(T)$ rather than $H_2(U)$ is valuable for improving welfare among households who are assigned to $d_1 = T$ by the first-period optimal policy $\pi_1^*$. We explore heterogeneity in the adaptive targeting effect in the next subsection.

## 5.3 Heterogeneity of Adaptive Targeting Effect

In the previous sections, we show that the adaptive targeting effect constitutes a major source of the welfare gains from dynamic targeting. This effect reflects the value of information, $H_2(T)$, for the planner in optimizing second-period treatment assignments and represents a key advantage of dynamic targeting relative to static targeting. In this section, we investigate which types of households exhibit larger adaptive targeting effects and how dynamic targeting leverages this heterogeneity in treatment allocation.

Figure 4 illustrates heterogeneity in the adaptive targeting effect across two observable household characteristics—past energy use and household income. Darker red indicates larger adaptive targeting effects, while darker blue indicates smaller effects. The figure reveals substantial heterogeneity in the adaptive targeting effect. For example, households with high past energy use and middle income (the right of the figure), as well as those with low past energy use and high income (the top-left of the figure), exhibit particularly large effects.

Our theory of the adaptive targeting effect in Section 2.3.3 predicts that when households in these groups—corresponding to the darker red cells in the figure—are treated in period 1, their treated outcomes generate especially valuable information for the planner in optimizing second-period treatment assignments.

To examine this mechanism empirically, we mark with black outlines the 15 cells with the highest ratios of *switchers*.[19] Switchers are households for which $\pi_1^*(H_1) = T$ but $\pi_{S,1}^*(H_1) = U$—that is, households that would remain untreated in period 1 under the static targeting policy but would receive treatment in period 1 under the dynamic targeting policy. We find that the adaptive targeting effect is particularly large in these outlined cells, suggesting that this mechanism plays an important role in shifting households from untreated under the static policy to treated under the dynamic policy in period 1.

# 6 Conclusion

Economic policies often involve dynamic interventions, in which individuals receive repeated treatments over multiple periods. In this study, we develop a framework for designing optimal dynamic targeting policies that maximize social welfare gains from policy interventions. We show that dynamic targeting can outperform static targeting through several key mechanisms—learning, habit formation, and adaptive targeting effects—and we develop a method to empirically identify these effects.

Using a sequential randomized controlled trial of a residential energy rebate program, we find that dynamic targeting nearly doubles the social welfare gains relative to static targeting and produces welfare gains more than five times as large as those under non-targeting policies. We also investigate how potential manipulation, if policy rules are revealed, would affect welfare gains from dynamic targeting. Our results demonstrate that dynamic targeting has the potential to substantially enhance welfare gains from a wide range of economic policies.

[19] These 15 cells account for roughly 10% of the experimental sample.

# References


ABBRING, J. H. AND G. J. VAN DEN BERG (2003): "The nonparametric identification of treatment effects in duration models," *Econometrica*, 71, 1491–1517.

ANDREWS, I., T. KITAGAWA, AND A. MCCLOSKEY (2024): "Inference on winners," *The Quarterly Journal of Economics*, 139, 305–358.

ANDREWS, I. S., T. KITAGAWA, AND A. MCCLOSKEY (2019): "Inference on Winners," *NBER working paper*.

ARIU, K., M. KATO, J. KOMIYAMA, K. MCALINN, AND C. QIN (2021): "Policy choice and best arm identification: Asymptotic analysis of exploration sampling," *arXiv preprint arXiv:2109.08229*.

ATHEY, S. AND E. PALIKOT (2026): "Effective and scalable programs to facilitate labor market transitions for women in technology," *arXiv*.

AVIN, C., I. SHPITSER, AND J. PEARL (2005): "Identifiability of path-specific effects," in *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*, 357–363.

BECKER, G. S. AND K. M. MURPHY (1988): "A Theory of Rational Addiction," *Journal of Political Economy*, 96, 675–700.

BREIMAN, L., J. H. FRIEDMAN, R. A. OLSHEN, AND C. J. STONE (2017): *Classification and regression trees*, Routledge.

BURLIG, F., C. KNITTEL, D. RAPSON, M. REGUANT, AND C. WOLFRAM (2020): "Machine learning from schools about energy efficiency," *Journal of the Association of Environmental and Resource Economists*, 7, 1181–1217.

CHAKRABORTY, B. AND E. E. M. MOODIE (2013): *Statistical Methods for Dynamic Treatment Regimes*, New York: Springer.

CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, AND B. MELLY (2013): "Inference on counterfactual distributions," *Econometrica*, 81, 2205–2268.

CHERNOZHUKOV, V. AND C. HANSEN (2005): "An IV model of quantile treatment effects," *Econometrica*, 73, 245–261.

DESHPANDE, M. AND Y. LI (2019): "Who Is Screened Out? Application Costs and the Targeting of Disability Programs," *American Economic Journal: Economic Policy*, 11, 213–248.

DIMAKOPOULOU, M., Z. ZHOU, S. ATHEY, AND G. IMBENS (2017): "Estimation considerations in contextual bandits," *arXiv preprint arXiv:1711.07077*.

DING, W. AND S. F. LEHRER (2010): "Estimating treatment effects from contaminated multiperiod education experiments: The dynamic impacts of class size reductions," *The Review of Economics and Statistics*, 92, 31–42.

FINKELSTEIN, A. AND M. J. NOTOWIDIGDO (2019): "Take-up and Targeting: Experimental Evidence from SNAP," *Quarterly Journal of Economics*, 134, 1505–1556.

FRIEDBERG, R., J. TIBSHIRANI, S. ATHEY, AND S. WAGER (2021): "Local Linear Forests," *Journal of Computational and Graphical Statistics*, 30, 503–517.

HASTIE, T., R. TIBSHIRANI, AND J. FRIEDMAN (2009): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, Springer New York, 2nd ed.

HECKMAN, J. J., H. ICHIMURA, AND P. E. TODD (1997): "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *The Review of Economic Studies*, 64, 605–654.

HIRANO, K. AND J. R. PORTER (2009): "Asymptotics for statistical treatment rules," *Econometrica*, 77, 1683–1701.

HUBER, M. (2019): "A review of causal mediation analysis for assessing direct and indirect treatment effects (SES Working Paper 500)," *Fribourg, Switzerland: University of Fribourg*.

IDA, T., T. ISHIHARA, K. ITO, D. KIDO, T. KITAGAWA, S. SAKAGUCHI, AND S. SASAKI (2026): "Choosing Who Chooses: Selection-driven targeting in energy rebate programs," *Econometrica*, 94, 225–247.

IDA, T., T. ISHIHARA, D. KIDO, AND S. SASAKI (2020): "A field experiment using rebates and machine learnings to promote energy-saving behavior," *AEA RCT Registry, No. 6139.*

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, 240–67.

——— (2023): "Selection on Welfare Gains: Experimental Evidence from Electricity Plan Choice," *American Economic Review*, 113, 2937–73.

KASY, M. AND A. SAUTMANN (2021): "Adaptive treatment assignment in experiments for policy choice," *Econometrica*, 89, 113–132.

KITAGAWA, T. AND A. TETENOV (2018): "Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice," *Econometrica*, 86, 591–616.

KNITTEL, C. R. AND S. STOLPER (2021): "Machine Learning about Treatment Effect Heterogeneity: The Case of Household Energy Use," *AEA Papers and Proceedings*, 111, 440–44.

KO, R., K. UETAKE, K. YATA, AND R. OKADA (2022): "When to Target Customers? Retention Management using Dynamic Off-Policy Policy Learning," *SSRN*.

KOCK, A. B., D. PREINERSTORFER, AND B. VELIYEV (2022): "Functional sequential treatment allocation," *Journal of the American Statistical Association*, 117, 1311–1323.

Kolsrud, J., C. Landais, P. Nilsson, and J. Spinnewijn (2018): "The optimal timing of unemployment benefits: Theory and evidence from Sweden," *American Economic Review*, 108, 985–1033.

Lattimore, T. and C. Szepesvári (2020): *Bandit algorithms*, Cambridge University Press.

Lechner, M. (2009): "Sequential causal models for the evaluation of labor market programs," *Journal of Business & Economic Statistics*, 27, 71–83.

Liu, X. (2023): "Dynamic Coupon Targeting Using Batch Deep Reinforcement Learning: An Application to Livestream Shopping," *Marketing Science*, 42, 637–658.

Luckett, D. J., E. B. Laber, A. R. Kahkoska, D. M. Maahs, E. Mayer-Davis, and M. R. Kosorok (2019): "Estimating dynamic treatment regimes in mobile health using v-learning," *Journal of the American Statistical Association*.

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

Meyer, B. D. (1995): "Lessons from the U.S. unemployment insurance experiments," *Journal of Economic Literature*, 33, 91–131.

Murphy, S. A. (2003): "Optimal dynamic treatment regimes," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65, 331–355.

Pelham Jr, W. E., G. A. Fabiano, J. G. Waxmonsky, A. R. Greiner, E. M. Gnagy, W. E. Pelham III, S. Coxe, J. Verley, I. Bhatia, K. Hart, et al. (2016): "Treatment sequencing for childhood ADHD: A multiple-randomization study of adaptive medication and behavioral interventions," *Journal of Clinical Child & Adolescent Psychology*, 45, 396–415.

Robins, J. M. (1986): "A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect," *Mathematical Modelling*, 7, 1393–1512.

Rodríguez, J., F. Saltiel, and S. Urzúa (2022): "Dynamic treatment effects of job training," *Journal of Applied Econometrics*, 37, 242–269.

Sakaguchi, S. (2025): "Estimation of optimal dynamic treatment assignment rules under policy constraints," *Quantitative Economics*, 16, 981–1022.

Tsiatis, A. A., M. Davidian, S. T. Holloway, and E. B. Laber (2019): *Dynamic Treatment Regimes: Statistical Methods for Precision Medicine*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, CRC Press.

Vuong, Q. and H. Xu (2017): "Counterfactual mapping and individual treatment effects in nonseparable models with binary endogeneity," *Quantitative Economics*, 8, 589–610.

Wager, S. and S. Athey (2018): "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests," *Journal of the American Statistical Association*, 113, 1228–1242.

WOLAK, F. A. (2011): “Do Residential Customers Respond to Hourly Prices? Evidence from a Dynamic Pricing Experiment,” *The American Economic Review*, 101, 83–87.

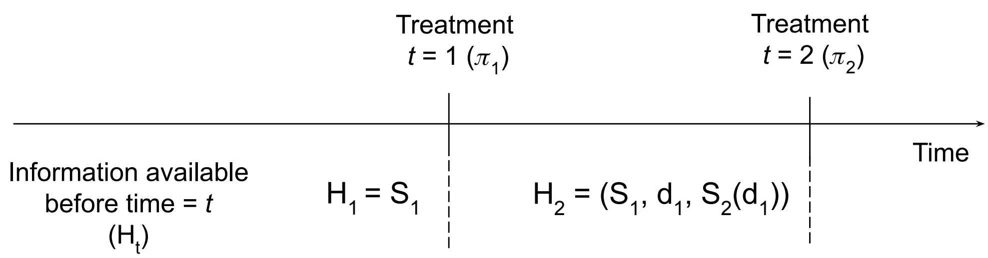
ZHANG, Y., E. B. LABER, M. DAVIDIAN, AND A. A. TSIATIS (2018): “Interpretable dynamic treatment regimes,” *Journal of the American Statistical Association*, 113, 1541–1549.

ZHAO, Y., D. ZENG, A. J. RUSH, AND M. R. KOSOROK (2012): “Estimating individualized treatment rules using outcome weighted learning,” *Journal of the American Statistical Association*, 107, 1106–1118.

ZHOU, Z., S. ATHEY, AND S. WAGER (2023): “Offline Multi-Action Policy Learning: Generalization and Optimization,” *Operations Research*, 71, 148–183.

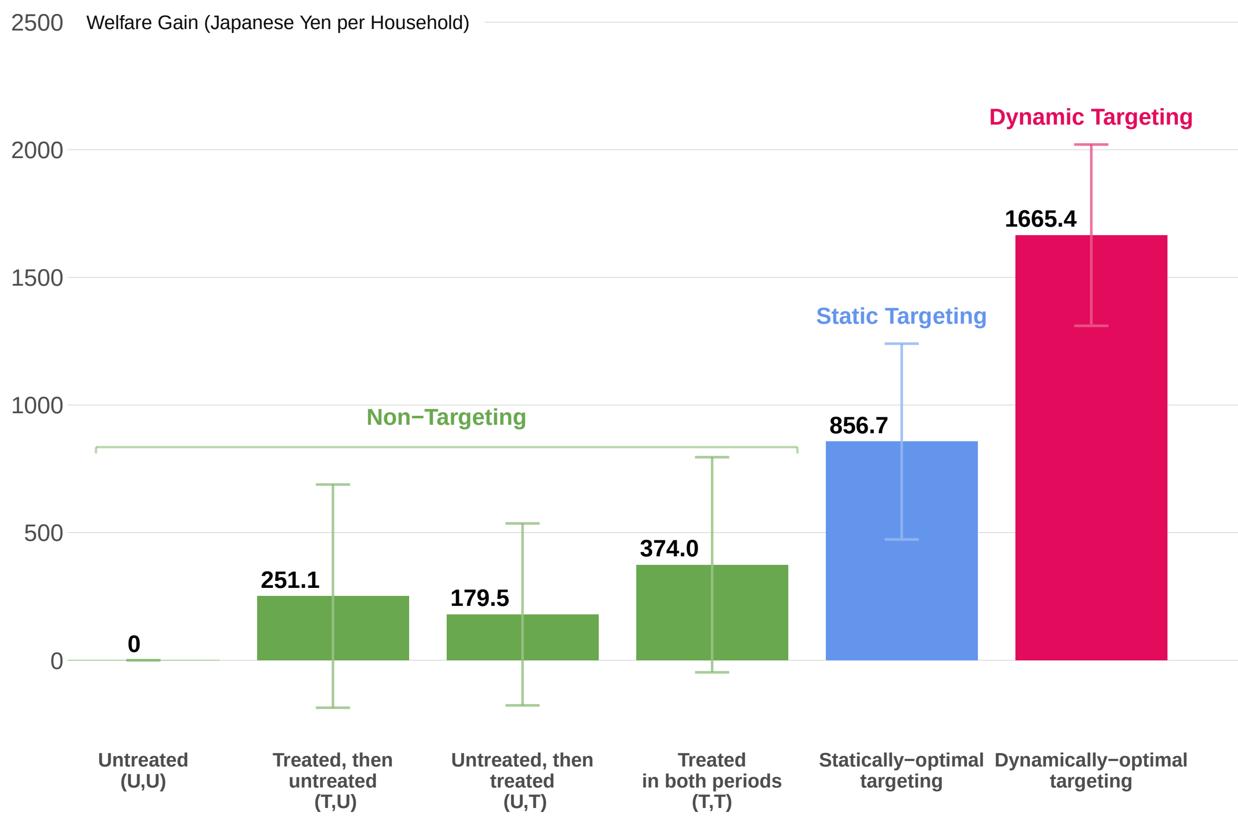
# Figures

Figure 1: Information Available to the Planner



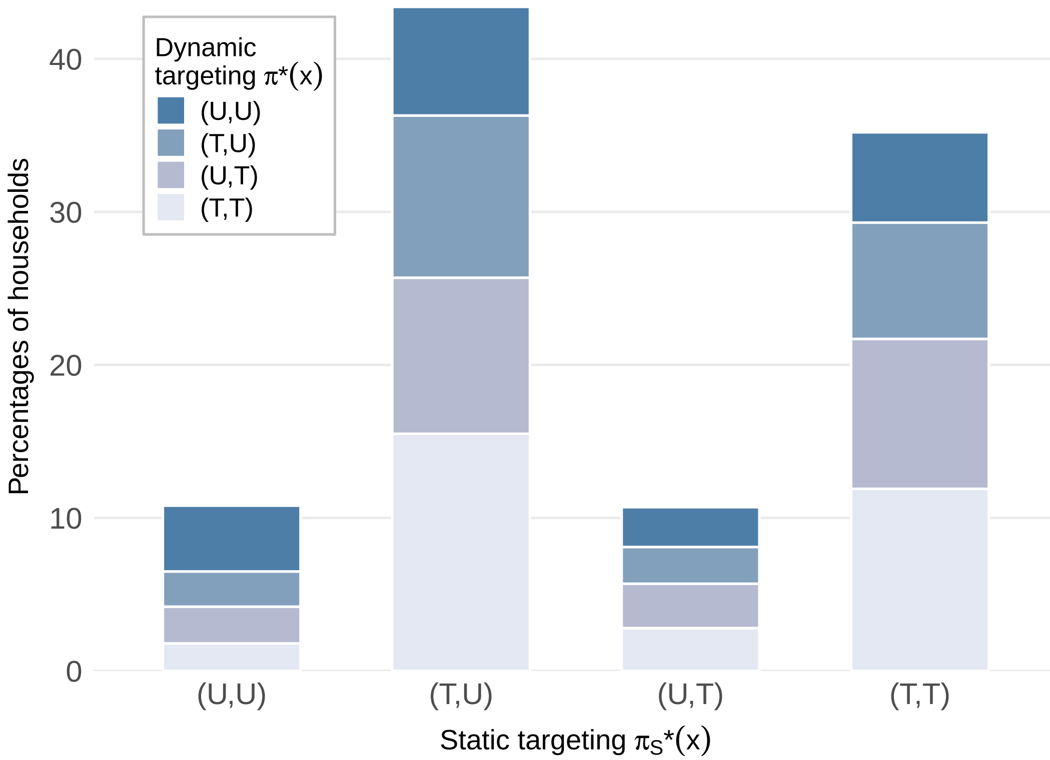

Notes: This figure depicts the information available to the planner prior to time $t$, denoted by $H_t$. At $t = 1$, $H_1$ agrees with $S_1$, and, at $t = 2$, $H_2(d_1) = (S_1, d_1, S_2(d_1))$, where $S_t$ is the observable data at time $t$, and $d_1$ is the treatment assignment at $t = 1$.

Figure 2: Welfare Gains from Each Policy



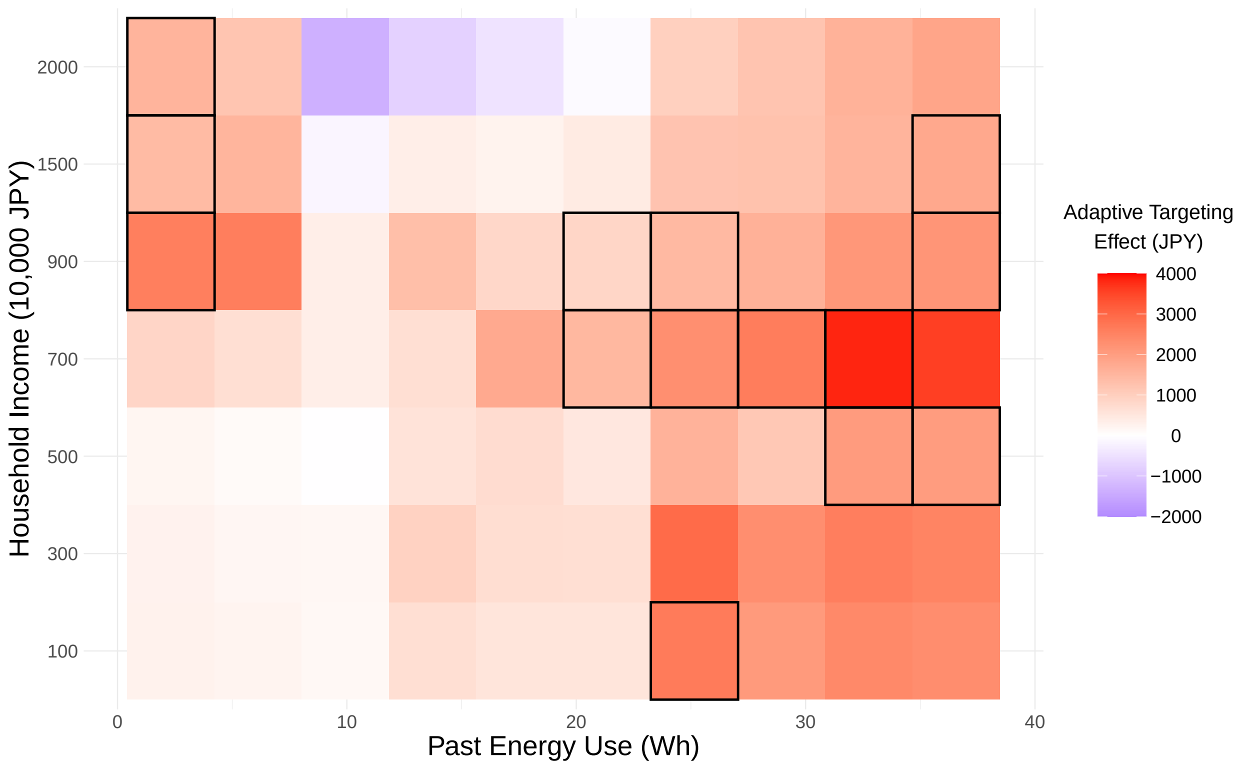

Notes: This figure shows the welfare gains of four non-targeting policies (100% $(U, U)$, 100% $(T, U)$, 100% $(U, T)$, and 100% $(T, T)$) and the optimal static and dynamic targeting policies. The welfare gain of each policy is represented by a bar, with estimated values displayed in JPY per household and error bars indicating 95 percent confidence intervals. The monetary unit is given as 1 ¢ $\approx$ 1 JPY, the exchange rate in 2020.

Figure 3: Differences in Optimal Assignment: Static vs. Dynamic Targeting




Notes: The height of each bar shows the percentage of households assigned to each of $((U,U),(T,U),(U,T),(T,T))$ under the optimal static targeting $(\pi_s^*)$. Each bar is further partitioned to reflect assignments under the optimal dynamic targeting rule $(\pi^*)$.

Figure 4: Heterogeneity in Adaptive Targeting Effect




Notes: This figure illustrates heterogeneity in the adaptive targeting effect across two observable household characteristics—past energy use and household income. Darker red indicates larger adaptive targeting effects, while darker blue indicates smaller effects. We mark with black outlines the 15 cells with the highest ratios of *switchers*. These top 15 cells account for roughly 10% of the experimental sample.

# Tables

Table 1: Summary Statistics and Covariate Balance Across Randomized Groups

| | Sample mean by group [standard deviation] | | | |
|---|---|---|---|---|
| | $(U,U)$ | $(U,T)$ | $(T,U)$ | $(T,T)$ |
| Peak hour usage (2020 summer, Wh) | 201 [145] | 200 [136] | 196 [136] | 198 [136] |
| Pre-peak hour usage (2020 summer, Wh) | 189 [143] | 184 [130] | 183 [137] | 182 [130] |
| Post-peak hour usage (2020 summer, Wh) | 311 [175] | 311 [171] | 308 [164] | 305 [163] |
| Peak hour usage (2020 winter, Wh) | 311 [194] | 309 [170] | 304 [179] | 306 [170] |
| Pre-peak hour usage (2020 winter, Wh) | 171 [117] | 171 [102] | 169 [112] | 166 [102] |
| Post-peak hour usage (2020 winter, Wh) | 287 [198] | 295 [198] | 280 [203] | 287 [192] |
| Number of people at home (1-5 PM) | 1.31 [1.04] | 1.32 [0.96] | 1.31 [1.04] | 1.34 [1.01] |
| Number of people at home (5-9 PM) | 2.57 [1.29] | 2.48 [1.20] | 2.47 [1.23] | 2.51 [1.20] |
| Self-efficacy in energy conservation (1-5 scale) | 3.44 [0.84] | 3.44 [0.86] | 3.47 [0.86] | 3.44 [0.82] |
| Household income (JPY 10,000) | 651 [400] | 639 [387] | 614 [393] | 606 [333] |

Notes: Columns 1-4 show present sample means and standard deviations in brackets for the pre-experimental consumption data and demographic variables by randomly-assigned group to the first and second period. $(U,U)$ is assigned to the untreated in both the first and second periods. $(U,T)$ is assigned to the untreated in the first period and the treated in the second period. $(T,U)$ is assigned to the treated in the first period and the untreated in the second period. $(T,T)$ is assigned to the treated in both the first and second periods. The number of households are 625 for $(U,U)$, 606 for $(U,T)$, 581 for $(T,U)$, and 588 for $(T,T)$. The monetary unit is given as 1 ¢ $\approx$ 1 JPY in 2020.

Table 2: Welfare Gains from Each Policy

| Policy | Welfare gain | Share of customers in each arm | | | |
|---|---|---|---|---|---|
| | | $(U, U)$ | $(T, U)$ | $(U, T)$ | $(T, T)$ |
| 100% $(U, U)$ | 0.0<br>(0.0) | 100.0% | 0.0% | 0.0% | 0.0% |
| 100% $(T, U)$ | 251.1<br>(437.3) | 0.0% | 100.0% | 0.0% | 0.0% |
| 100% $(U, T)$ | 179.5<br>(356.5) | 0.0% | 0.0% | 100.0% | 0.0% |
| 100% $(T, T)$ | 374.0<br>(421.5) | 0.0% | 0.0% | 0.0% | 100.0% |
| Static targeting ($\pi_S^*$) | 856.7<br>(383.6) | 10.7% | 43.5% | 10.7% | 35.2% |
| Dynamic targeting ($\pi^*$) | 1665.4<br>(335.2) | 19.9% | 22.9% | 25.2% | 32.0% |

Notes: This table summarizes characteristics of four benchmark policies (100% $(U, U)$, 100% $(T, U)$, 100% $(U, T)$, and 100% $(T, T)$), static targeting ($\pi_S^*$), and dynamic targeting ($\pi^*$). The column titled "Welfare Gain" shows the estimated welfare gains in JPY per household, with these standard errors in parentheses. The monetary unit is given as 1 ¢ $\approx$ 1 JPY, the exchange rate in 2020.

Table 3: Welfare Differences Between Policies

| | Welfare difference | | |
|---|---|---|---|
| | Period 1 | Period 2 | Total |
| Dynamic targeting ($\pi^*$) vs. 100% $(U, U)$ | 348.7 | 1316.7 | 1665.4 |
| | (110.0) | (287.8) | (335.2) |
| Dynamic targeting ($\pi^*$) vs. 100% $(T, U)$ | 103.5 | 1310.9 | 1414.3 |
| | (102.3) | (277.6) | (319.4) |
| Dynamic targeting ($\pi^*$) vs. 100% $(U, T)$ | 348.7 | 1137.2 | 1485.9 |
| | (110.0) | (298.4) | (343.4) |
| Dynamic targeting ($\pi^*$) vs. 100% $(T, T)$ | 103.5 | 1187.9 | 1291.4 |
| | (102.3) | (288.9) | (329.9) |
| Dynamic targeting ($\pi^*$) vs. Static targeting ($\pi_S^*$) | $-280.9$ | 1089.6 | 808.7 |
| | (90.8) | (266.2) | (302.4) |

Notes: This table compares the welfare gains from each policy. For each row, the columns "Period 1" and "Period 2" report the estimated welfare gains in periods 1 and 2, respectively, measured in JPY per household per season, with standard errors in parentheses. The monetary unit is expressed such that 1 ¢ $\approx$ 1 JPY, corresponding to the exchange rate in 2020.

Table 4: Potential Anticipatory Responses and Welfare Implications

| Scenario | Welfare gain | Share of customers in each arm | | | |
|---|---|---|---|---|---|
| | | $(U, U)$ | $(T, U)$ | $(U, T)$ | $(T, T)$ |
| No manipulation | 1665.4 | 19.9% | 22.9% | 25.2% | 32.0% |
| | (335.2) | | | | |
| Downward manipulation | 1666.2 | 18.5% | 21.0% | 26.6% | 33.9% |
| | (336.7) | | | | |
| Upward manipulation | 1650.7 | 19.8% | 22.8% | 25.3% | 32.2% |
| | (335.4) | | | | |
| Both manipulations | 1651.0 | 18.4% | 20.9% | 26.7% | 34.0% |
| | (336.9) | | | | |

Notes: This table reports estimated welfare gains and the corresponding share of households assigned to each arm under different manipulation scenarios discussed in Section 4.4. Welfare gains are measured in JPY per household, with standard errors in parentheses. The monetary unit is expressed such that 1 ¢ $\approx$ 1 JPY, corresponding to the exchange rate in 2020.

Table 5: Decomposition of Welfare Gains from the Dynamic Targeting $\pi^*$

| | Conditional effect | Fraction | Welfare contribution |
|---|---|---|---|
| 1st-period treatment effect | 493.7 | 0.55 | 271.1 |
| | (199.9) | (0.00) | (109.8) |
| 2nd-period treatment effect | 1432.7 | 0.56 | 801.7 |
| | (386.4) | (0.00) | (215.9) |
| Habit formation effect | 911.6 | 0.22 | 204.7 |
| | (968.3) | (0.01) | (204.6) |
| Learning effect | 22.9 | 0.32 | 7.4 |
| | (407.5) | (0.01) | (126.6) |
| Adaptive targeting effect | 668.3 | 0.55 | 367.0 |
| | (199.8) | (0.01) | (109.7) |
| Total effect | | | 1652.0 |
| | | | (442.8) |

Notes: This table shows estimation results for the decomposition (5). The column "Conditional effect" shows the estimates of conditional average effects that appear in equation (5), and the column "Fraction" shows each estimate of the fraction of each conditioning group. The column "Welfare contribution" shows each estimate of each term in the right hand side of equation (5). For example, regarding the row "Learning effect", the column "Conditional effect" shows the estimate of $E[Y_2(T,T) - Y_2(U,T)|\pi_1^*, (H_1) = T, \pi_2^*(H_2(T)) = T]$, the column "Fraction" shows the estimate of $\Pr(\pi_1^*, (H_1) = T, \pi_2^*(H_2(T)) = T)$, and the column "Welfare contribution" shows the estimate of $E[Y_2(T,T) - Y_2(U,T)|\pi_1^*, (H_1) = T, \pi_2^*(H_2(T)) = T] \times \Pr(\pi_1^*, (H_1) = T, \pi_2^*(H_2(T)) = T)$. The last row "Total effect" shows the estimate of the sum of all terms in the right hand side of equation (5). The standard errors are in parentheses. The monetary unit is given as 1 ¢ ≈ 1 JPY, the exchange rate in 2020.

# A Online Appendix

## A.1 Estimation of the Welfare Gain Decomposition

This appendix describes the estimation of three components in the decomposition (5): the conditional averages of the habit-formation effect, the learning effect, and the adaptive targeting effect. The proposed estimation builds on the identification results in Proposition 5.1.

Let $\widehat{F}_{S_2^k|(D_1,S_1)}(\cdot \mid d_1, s_1)$ and $\widehat{Q}_{S_2^k|(D_1,S_1)}(\cdot \mid d_1, s_1)$ denote estimators of the conditional distribution function $F_{S_2^k|(D_1,S_1)}(\cdot \mid d_1, s_1)$ and the conditional quantile function $Q_{S_2^k|(D_1,S_1)}(\cdot \mid d_1, s_1)$, respectively. To estimate $F_{S_2^k|(D_1,S_1)}(\cdot \mid d_1, s_1)$, we apply distribution regression with the standard logistic distribution function (e.g., Chernozhukov et al., 2013), using the subsample with $D_1 = d_1$. To estimate $Q_{S_2^k|(D_1,S_1)}(\cdot \mid d_1, s_1)$, we use linear quantile regression on the subsample with $D_1 = d_1$. Both estimations are implemented using the constructed test data.

Using the test data, we estimate the counterfactual state variables $S_{i2}^k(d_1)$ for each unit $i$ as

$$\widehat{S}_{i2}^{k,\text{test}}(d_1) = \mathbf{1}\{D_{i1}^{\text{test}} = d_1\} \cdot S_{i2}^{k,\text{test}} + \mathbf{1}\{D_{i1}^{\text{test}} \neq d_1\} \cdot \widehat{Q}_{S_2^k|(D_1,S_1)}\left(\widehat{F}_{S_2^k|(D_1,S_1)}(S_{i1}^{k,\text{test}}|D_{i1}^{\text{test}}, S_{i1}^{\text{test}})\middle| d_1, S_{i1}^{\text{test}}\right).$$

Define $\widehat{S}_{i2}^{\text{test}}(d_1) = (\widehat{S}_{i2}^{1,\text{test}}, \widehat{S}_{i2}^{2,\text{test}})$ and $\widehat{H}_{i2}^{\text{test}}(d_1) = \left(S_{i1}^{\text{test}}, d_1, \widehat{S}_{i2}^{\text{test}}(d_1)\right)$.

Based on these constructions, we estimate the three counterfactual components of the decomposition—the average habit-formation effect, learning effect, and adaptive targeting effect—as follows:

- Conditional average habit-formation effect (units assigned to $(T, U)$):

$$\sum_{i=1}^{n}\left[\frac{\mathbf{1}\{\hat{\pi}_1^*(H_{i1}^{\text{test}}) = T, \hat{\pi}_2^*\left(\widehat{H}_{i2}^{\text{test}}(T)\right) = U\}}{\sum_{i=1}^{n}\mathbf{1}\{\hat{\pi}_1^*(H_{i1}^{\text{test}}) = T, \hat{\pi}_2^*\left(\widehat{H}_{i2}^{\text{test}}(T)\right) = U\}} \right.$$
$$\left. \times \left(\frac{\mathbf{1}\{D_{i1}^{\text{test}} = T\}}{P(D_1 = T|H_1 = H_{i1}^{\text{test}})} - \frac{\mathbf{1}\{D_{i1}^{\text{test}} = U\}}{P(D_1 = U|H_1 = H_{i1}^{\text{test}})}\right) \cdot \frac{\mathbf{1}\{D_{i2}^{\text{test}} = U\}}{P(D_2 = U|H_2 = H_{i2}^{\text{test}})} \cdot Y_{i2}^{\text{test}}\right].$$

- Conditional average learning effect (units assigned to $(T, T)$):

$$\sum_{i=1}^{n}\left[\frac{\mathbf{1}\{\hat{\pi}_1^*(H_{i1}^{\text{test}})=T, \hat{\pi}_2^*\left(\widehat{H}_{i2}^{\text{test}}(T)\right)=T\}}{\sum_{i=1}^{n}\mathbf{1}\{\hat{\pi}_1^*(H_{i1}^{\text{test}})=T, \hat{\pi}_2^*\left(\widehat{H}_{i2}^{\text{test}}(T)\right)=T\}}\right.$$
$$\left.\times\left(\frac{\mathbf{1}\{D_{i1}^{\text{test}}=T\}}{P(D_1=T|H_1=H_{i1}^{\text{test}})}-\frac{\mathbf{1}\{D_{i1}^{\text{test}}=U\}}{P(D_1=U|H_1=H_{i1}^{\text{test}})}\right)\cdot\frac{\mathbf{1}\{D_{i2}^{\text{test}}=T\}}{P(D_2=T|H_2=H_{i2}^{\text{test}})}\cdot Y_{i2}^{\text{test}}\right]:$$

- Conditional average adaptive targeting effect (units treated in the first period):

$$\sum_{i=1}^{n}\left[\frac{\mathbf{1}\{\hat{\pi}_1^*(H_{i1}^{\text{test}})=T\}}{\sum_{i=1}^{n}\mathbf{1}\{\hat{\pi}_1^*(H_{i1}^{\text{test}})=T\}}\cdot\frac{\mathbf{1}\{D_{i1}^{\text{test}}=T\}}{P(D_1=T|H_1=H_{i1}^{\text{test}})}\right.$$
$$\left.\times\left(\frac{\mathbf{1}\left\{D_{i2}^{\text{test}}=\hat{\pi}_2^*\left(\widehat{H}_{i2}^{\text{test}}(T)\right)\right\}}{P\left(D_2=\hat{\pi}_2^*\left(\widehat{H}_{i2}^{\text{test}}(T)\right)\middle|H_2=H_{i2}^{\text{test}}\right)}-\frac{\mathbf{1}\left\{D_{i2}^{\text{test}}=\hat{\pi}_2^*\left(\widehat{H}_{i2}^{\text{test}}(U)\right)\right\}}{P\left(D_2=\hat{\pi}_2^*\left(\widehat{H}_{i2}^{\text{test}}(U)\right)\middle|H_2=H_{i2}^{\text{test}}\right)}\right)\cdot Y_{i2}^{\text{test}}\right].$$

## A.2 Proof of Proposition 2.1

We prove a general version of the proposition by letting the dynamic targeting policy $\pi$ be arbitrary.

Let $W(d_1) = E[Y_1(d_1)]$ and $W_2(d_1, d_2) = E[Y_2(d_1, d_2)]$. Consider first the following decomposition, which follows from the law of total probability with respect to the event $\pi_1(H_1) = U$ and its complement:

$$\begin{aligned}
&W(\pi) - W_{UU} \\
&= E[Y_2(U, \pi_2(H_2(U))) - Y_2(U, U)|\pi_1(H_1) = U]\Pr(\pi_1(H_1) = U) && (10)\\
&+ E[Y_1(T) - Y_1(U)|\pi_1(H_1) = T]\Pr(\pi_1(H_1) = T) && (11)\\
&+ E[Y_2(T, \pi_2(H_2(T))) - Y_2(U, U)|\pi_1(H_1) = T]\Pr(\pi_1(H_1) = T). && (12)
\end{aligned}$$

Given a DTR $\pi$, we define the following subsets of the population:

$$A_1(U) = \{\pi_1(H_1) = U\};\ A_1(T) = \{\pi_1(H_1) = T\};$$
$$A_2(U) = \{\pi_2(H_2(U)) = U\};\ A_2(T) = \{\pi_2(H_2(U)) = T\}.$$

Note that Term (10) is zero if $\pi_2(H_2(U)) = U$. Hence, it can be expressed as

$$(10) = E[W_2(U,T) - W_2(U,U)|A_2(T) \cap A_1(U)] \cdot \Pr(A_2(T) \cap A_1(U)). \tag{13}$$

Term (12) can be decomposed as

$$(12) = E[Y_2(T,\pi_2(H_2(T))) - Y_2(T,\pi_2(H_2(U)))|A_1(T)] \Pr(A_1(T)) \tag{14}$$
$$+ E[Y_2(T,\pi_2(H_2(U))) - Y_2(U,\pi_2(H_2(U)))|A_1(T)] \Pr(A_1(T)) \tag{15}$$
$$+ E[Y_2(U,\pi_2(H_2(U))) - Y_2(U,U)|A_1(T)] \Pr(A_1(T)). \tag{16}$$

Similar to (13), Term (16) is zero if $\pi_2(H_2(U)) = U$. Therefore, we obtain

$$(16) = E[Y_2(U,T) - Y_2(U,U)|A_2(T) \cap A_1(T)] \cdot \Pr(A_2(T) \cap A_1(U)). \tag{17}$$

The sum of terms (13) and (17) yields

$$(13) + (16) = E[Y_2(U,T) - Y_2(U,U)|A_2(T)] \cdot \Pr(A_2(T)). \tag{18}$$

Term (15) can be expressed as

$$(15) = E[Y_2(T,U) - Y_2(U,U)|A_2(U) \cap A_1(T)] \Pr(A_2(U) \cap A_1(T))$$
$$+ E[Y_2(T,T) - Y_2(U,T)|A_2(T) \cap A_1(T)] \Pr(A_2(T) \cap A_1(T)). \tag{19}$$

By combining Terms (11), (14), (18), and (19), we obtain the proposition. $\square$

## A.3 Welfare Gain Decomposition for Static Targeting

We can also decompose the welfare gain of the optimal static targeting $\pi_S^*$ in a manner similar to that of the optimal dynamic targeting. Specifically, it can be decomposed into four components:[20]

$$
\begin{aligned}
W(\pi_S^*) - W_{UU} = & \underbrace{E[Y_1(T) - Y_1(U)|\pi_{S,1}^*(H_1) = T]}_{\text{Treatment effect on the treated in } t=1} \cdot \Pr(\pi_{S,1}^*(H_1) = T) \\
& + \underbrace{E[Y_2(U,T) - Y_2(U,U)|\pi_{S,2}^*(H_1) = T]}_{\text{Treatment effect on the treated in } t=2} \cdot \Pr(\pi_{S,2}^*(H_1) = T) \\
& + \underbrace{E[Y_2(T,U) - Y_2(U,U)|\pi_S^*(H_1) = (T,U)]}_{\text{Habit formation effect for those assigned to } (T,U)} \cdot \Pr(\pi_S^*(H_1) = (T,U)) \\
& + \underbrace{E[Y_2(T,T) - Y_2(U,T)|\pi_S^*(H_1) = (T,T)]}_{\text{Learning effect for those assigned to } (T,T)} \cdot \Pr(\pi_S^*(H_1) = (T,T)).
\end{aligned}
$$

Importantly, the welfare gain decomposition for static targeting does not include the adaptive targeting effect. Intuitively, this is because static targeting does not use the updated information $H_2(d_1)$ when making the second-period treatment choice. This result highlights that only dynamic targeting exploits the adaptive targeting effect to improve welfare.

## A.4 The External Validity of the Experimental Sample

We randomly sampled 2070 customers from the target population who did not participate in this experiment, and conducted a similar survey to the one for the experimental sample. The purpose of this was to investigate the external validity of our experimental sample by comparing the mean for each variable between the control group from our experimental sample and this random sample. Columns 1 and 2 of Table A.1 present summary statistics for the untreated group and the random sample. Column 3 presents differences in means, with the standard errors of these differences in parentheses. We observe larger means for four variables in the untreated group than in the random sample, and the differences are statistically significant. Our experimental sample has

[20]The decomposition result for $\pi_S^*$ follows directly by replacing $\pi^*$ with $\pi_S^*$ in the proof of Proposition 2.1.

larger pre-experiment electricity usage per month, a larger number of people at home on weekdays, higher self-efficacy in energy conservation, and higher household income. This implies that our sample includes a larger number of customers who are willing and able to reduce their electricity consumption, which should be taken into consideration when discussing the generalizability of this study.

[Table A.1 about here]

## A.5 Exogenous Parameters in Social Welfare Criterion

Our social welfare criterion depends on four exogenous parameters: $p$, $c$, $a$, and $\delta$. We calibrate these parameters using data from the Japanese electricity market during our experimental period. The parameter $p$ denotes the unit retail price of electricity. We set $p = 25$ JPY/kWh, corresponding to the regulated price of electricity in Japan, which is independent of the time of day.[21] The parameter $c$ represents the marginal cost of electricity production. We set $c = 125$ JPY/kWh so that the cost–price difference, $c - p$, equals the rebate amount of 100 JPY per kWh. The wholesale price of electricity often spikes during peak hours, such as summer afternoons or winter evenings, due to supply constraints. Historically, wholesale prices have occasionally exceeded 100 JPY/kWh.[22]

The parameter $a$ represents the administrative cost of implementing the energy-saving program. This cost includes several items, such as the installation cost of the Home Energy Management System (HEMS) required for participation. In 2016, the Japanese government estimated the implementation cost of a demand reduction program, including HEMS installation, at 291.1 JPY per household per season.[23] Because our analysis considers two periods, we set the administrative cost

[21] In Japan, until April 1, 2016, household electricity was supplied by local power companies at regulated retail prices. Since then, the retail electricity market has been fully liberalized, allowing households to freely choose their pricing plans. However, as a transitional measure, the regulated price for households has been maintained, at approximately 25 JPY/kWh regardless of the time of day.

[22] The wholesale electricity market, where generators and retailers trade electricity, is operated by the Japan Electric Power Exchange (JEPX). Most trading occurs in the "day-ahead market," where electricity is traded the day before the demand period. Trading results are published, and we confirm that the price exceeded 100 JPY/kWh on July 25, 2018. Moreover, prices have sometimes exceeded 125 JPY/kWh; for instance, the price reached 250 JPY/kWh on January 15, 2021.

[23] We do not include the installation cost of smart meters in the administrative cost. Since the Great East Japan

per period to half of this amount (291.1/2 JPY).

The parameter $\delta$ represents the long-term benefit of a unit reduction in energy consumption. Specifically, we consider the effect of such a reduction on the capacity market, where future supply capacity is traded between the generation and retail sectors. In Japan, the capacity market was established in 2020, with the first auction held in that year. In that auction, the Japanese government

## A.6 Linear Regression Analysis of Average Treatment Effects

The rebate program aims to incentivize energy conservation during peak hours; therefore, a key variable in our social welfare function is electricity usage during peak hours, which we present in Section 4.1. This section provides a regression analysis of the average treatment effect of the rebate program on peak-hour electricity usage in periods 1 and 2.

We evaluate the average treatment effect (ATE) of the randomly-assigned group $D_t = T$ relative to $D_t = U$ in each period ($t = 1, 2$) using ordinary least squares (OLS) with the following estimating equation:

$$Q_{ih} = \beta_T T_{ih} + \lambda_i + \theta_h + \epsilon_{ih}, \tag{20}$$

where $Q_{ih}$ is the natural log of electricity usage for household $i$ in a 30-minute interval $h$ in each period. We include data from the pre-experimental and experimental periods.[24] The dummy variable $T_{ih}$ equals one if household $i$ is in group $T$ and $h$ is in the treatment period. We include household fixed effects $\lambda_i$ and time fixed effects $\theta_h$ in each 30-minute interval to control for time-specific shocks such as weather. Given that $D_t = \{U, T\}$ is randomly assigned, $\beta_T$ provides the ATE of $D_t = T$ relative to $D_t = U$. We use cluster standard errors at the household level.

Earthquake of March 11, 2011, and the accident at the Fukushima Daiichi Nuclear Power Plant, the Japanese government has mandated the nationwide installation of smart meters in all homes by the end of the decade. This cost is therefore considered sunk, as it will be incurred regardless of whether a demand reduction program is implemented.

[24] Because of randomization, the pre-experimental data is not necessary for obtaining the consistent estimator. The primary benefit of including the pre-experimental data is that the inclusion of household fixed effects can substantially increase the precision of the estimates because residential electricity usage tends to form a significant part of household-specific time-invariant variation.

Table A.2 presents the estimation results for equation 20. The treatment reduced peak-hour electricity usage by 0.051 log points (5.0%) in the summer of 2020 and by 0.032 log points (3.2%) in winter. We confirm a significant average treatment effect for electricity savings in each period.

[Table A.2 about here]

Next, we examine the dynamic heterogeneity regarding the treatment effects on electricity usage in the second period. Specifically, we estimated the average treatment effect (ATE) with a baseline of $(D_1, D_2) = (U, U)$ for the randomly-assigned treatment groups $(D_1, D_2) = \{(U, T), (T, U), (T, T)\}$, using OLS with the following estimating equation:

$$Q_{ih} = \beta_{UT} UT_{ih} + \beta_{TU} TU_{ih} + \beta_{TT} TT_{ih} + \lambda_i + \theta_h + \epsilon_{ih}, \tag{21}$$

where $Q_{ih}$ is the natural log of electricity usage for household $i$ in a 30-minute interval $h$ in the winter of 2020 ($t = 2$). The dummy variable $UT_{ih}$ equals one if household $i$ is in group $U$ in the first period, group $T$ is in the second period, and $h$ is in the treatment period. Similarly, the dummy variable $TU_{ih}$ equals one if household $i$ is in group $T$ in the first period and group $U$ in the second period. The dummy variable $TT_{ih}$ equals one if household $i$ is in Group $T$ in the first and second periods. The coefficient $\beta_{TU}$ represents the extent to which the effect of the treatment received in the first period is sustained in the second period. Thus, this effect can be regarded as a habit-formation effect. We also regard a negative difference between the coefficients $\beta_{TT}$ and $\beta_{UT}$ as a learning effect and a positive difference as a habituation effect.

[Table A.3 about here]

Table A.3 presents the estimation results for equation (21). We begin by demonstrating the ATE for the entire sample in column 1. The treatment in the second period only $(U, T)$ reduced the peak-hour electricity usage by 0.046 log points (4.5%) in the second period. When the treatment is given only in the first period $(T, U)$, peak-hour electricity usage is reduced by 0.011 log points (1.1%), though this effect is not statistically significant. That is, the habit-formation effect was not observed on average ($p = 0.569$). The two treatments in the first and second periods $(T, T)$ reduced peak electricity usage by 0.029 log points (2.8%); however, this effect is not statistically significant. There is no statistically significant difference among $(T, T)$ and $(U, T)$ ($p = 0.351$). This result indicates neither the learning effect, in which the presence or absence of treatment in the first period positively affects the treatment effect in the second period in absolute terms, nor the habituation effect, in which the presence or absence of treatment in the first period negatively affects the treatment effect in the second period.

Beyond these overall program effects, an important question for our analysis is whether dynamic heterogeneity exists in these effects. If different types of households have different learning and habituation effects, the optimal dynamic targeting presented in Section 2 may increase welfare gains from the policy. The remaining columns of Table A.3 explore this issue. Each pair of columns divides the sample into two groups: those with a below-median value for a particular variable and those with an above-median value.

We find some evidence of heterogeneity in the program effects. In Columns 4 and 5, we divide customers by the difference in electricity usage between peak-hour and post-peak-hour at the summer baseline. For households with below-median values of this variable, we find that $\hat{\beta}_{UT} = -0.027$ and $\hat{\beta}_{TT} = -0.044$, and the p-value for the difference is 0.496. However, households with above-median values have $\hat{\beta}_{UT} = -0.070$ and $\hat{\beta}_{TT} = -0.013$, and the p-value for the difference is 0.037. That is, among the high group, $(U, T)$ induces a greater reduction than $(T, T)$, and the habituation effect is supported, though not for the low group. We find similar heterogeneity when we divide the sample by the number of people at home (1-5 PM) and self-efficacy for energy conservation. These heterogeneities in the effects of dynamically providing treatments on peak-hour electricity usage imply that optimal dynamic targeting is likely to enhance social welfare gains.

## A.7 Artificial Test Data

We describe our method for generating artificial test data in detail, building on Ida et al. (2026). To motivate this construction, we first explain why bias in point estimates of welfare gain arises when the same data are used both to learn an optimal policy and to estimate its welfare gain. Simply, bias results from noise in the observed electricity consumption $Q_1$ and $Q_2$, which are random components of $Y_1$ and $Y_2$. Specifically, the observed electricity consumption $Q_1$ and $Q_2$ can be decomposed as the sum of an essential term and noise as follows:

$$\begin{aligned}
Q_1 &= \underbrace{\sum_{d_1 \in \{T,U\}} E[Q_1(d_1)|H_1] \cdot \mathbf{1}\{D_1 = d_1\}}_{\text{essential term}} + \underbrace{\sum_{d_1 \in \{T,U\}} \epsilon_{1,d_1} \cdot \mathbf{1}\{D_1 = d_1\}}_{\text{noise term}}; \\
Q_2 &= \underbrace{\sum_{(d_1,d_2) \in \{T,U\}^2} E[Q_2(d_1,d_2)|H_2] \cdot \mathbf{1}\{(D_1,D_2) = (d_1,d_2)\}}_{\text{essential term}} \\
&\quad + \underbrace{\sum_{(d_1,d_2) \in \{T,U\}^2} \epsilon_{2,(d_1,d_2)} \cdot \mathbf{1}\{(D_1,D_2) = (d_1,d_2)\}}_{\text{noise term}},
\end{aligned}$$

where $\epsilon_{1,d_1} := Q_1(d_1) - E[Q_1(d_1)|H_1]$ and $\epsilon_{2,(d_1,d_2)} := Q_2(d_1,d_2) - E[Q_2(d_1,d_2)|H_2]$ for each $(d_1,d_2) \in \{T,U\}^2$. While only the essential term is relevant for learning an optimal policy, a learning algorithm inevitably responds to the noise term and thus overfits the training sample. Consequently, when the welfare performance of the estimated policy is evaluated on the same sample, the estimate is biased upward because the policy also fits the noise. This is known as the winner's curse bias (see., e.g., Andrews, Kitagawa, and McCloskey (2024)). Replacing the noise term in the training sample with an independent draw of noise removes this bias from the welfare estimate.

Motivated by this observation, we create an artificial test sample to mitigate the winner's curse bias in our welfare estimation. Specifically, for each period, we run random forest regressions of the outcome on the corresponding history with cross-fitting, and generate outcome observations as the sum of the regression fits and permuted residuals. We then estimate the optimal welfare by evaluating the welfare value in the test sample at the EWM-optimal policy obtained from the original sample.

Specifically, we construct test data $\mathcal{S}^{\text{test}} = \{S_{i1}, S_{i2}, Y_{i1}^{\text{test}}, Y_{i2}^{\text{test}}, D_{i1}, D_{i2}\}_{i=1}^{n}$ by generating $Q_{it}^{test}$ $(t = 1, 2)$ in the following procedure and substituting them into $Y_{it}^{\text{test}} = b \cdot Q_{it}^{\text{test}} - a \cdot \mathbf{1}\{D_{it} = T\}$. For each $(d_1, d_2) \in \{T, U\}^2$, let $I_{(d_1,d_2)} := \{i : (D_{i1}, D_{i2}) = (d_1, d_2)\}$ denote the set of observations assigned to $(d_1, d_2)$. We randomly partition $I_{(d_1,d_2)}$ into 10 equal-sized folds $I_{(d_1,d_2)}^{(1)}, \ldots, I_{(d_1,d_2)}^{(10)}$, and define $I_{(d_1,d_2)}^{(-k)} := I_{(d_1,d_2)}^{(-k)} \backslash I_{(d_1,d_2)}^{(k)}$, for $k = 1, \ldots, 10$.

For $Q_{i1}^{\text{test}}$ and $Q_{i2}^{\text{test}}$, we generate artificial data using the following procedure:

1. For each $(d_1, d_2) \in \{U, T\}^2$ and $k = 1, \ldots, 10$:

   (a) Estimate the conditional expectation function of the first-period electricity usage given intervention $d_1$ and history $H_1$, $E[Q_1(d_1)|H_1 = h_1]$ using observations in $I_{(d_1,d_2)}^{(-k)}$ to obtain $\widehat{E}^{(-k)}[Q_1(d_1)|H_1 = h_1]$. Next, compute the residuals $\hat{\epsilon}_{i1,d_1} = Q_{i1} - \hat{E}^{(-k)}[Q_1(d_1)|H_1 = H_{i1}]$ for $i \in I_{(d_1,d_2)}^{(k)}$. Subsequently, estimate the conditional variance of the error term $E[\epsilon_{i1,d_1}^2|H_1 = H_{i1}]$ by regressing $\hat{\epsilon}_{1i,d_1}^2$ on $H_1$ using observations from $I_{(d_1,d_2)}^{(-k)}$. Denote this estimator by $\hat{E}^{(-k)}[\epsilon_{i1,d_1}^2|H_1 = h_1]$, and compute $\hat{\sigma}_{i1,d_1}^2 = \hat{E}^{(-k)}[\epsilon_{1,d_1}^2|H_1 = H_{i1}]$.

   (b) Estimate the conditional expectation function of the potential second-period electricity usage, $E[Q_2(d_1, d_2)|H_2 = h_2]$, using observations in $I_{(d_1,d_2)}^{(-k)}$ to obtain $\hat{E}^{(-k)}[Q_2(d_1, d_2)|H_2 = H_{i2}]$. For $i \in I_{(d_1,d_2)}^{(k)}$, compute residuals $\hat{\epsilon}_{i2,(d_1,d_2)} = Q_{i2} - \hat{E}[Q_2(d_1, d_2)|H_2 = H_{i2}]$. Next, estimate the conditional variance of the error term $E[\epsilon_{i2,(d_1,d_2)}^2|H_2 = h_2]$ by regressing $\hat{\epsilon}_{i2,(d_1,d_2)}^2$ on $H_2$ using observations from $I_{(d_1,d_2)}^{(-k)}$, yielding $\hat{E}^{(-k)}[\epsilon_{i1,(d_1,d_2)}^2|H_2 = h_2]$. Finally, compute $\hat{\sigma}_{i2,(d_1,d_2)}^2 = \hat{E}^{(-k)}[\epsilon_{2,(d_1,d_2)}^2|H_2 = H_{i2}]$ $i \in I_{(d_1,d_2)}^{(k)}$.

   (c) Estimate the conditional covariance of the regression residuals, $E[\epsilon_{i1,d_1}\epsilon_{i2,(d_1,d_2)}|H_1 = H_{i1}]$, by regressing the product $\hat{\epsilon}_{i1,d_1} \cdot \hat{\epsilon}_{i2,(d_1,d_2)}$ on $H_1$ using observations in $I_{(d_1,d_2)}^{(-k)}$. This yields $\hat{E}^{(-k)}[\epsilon_{1,d_1}\epsilon_{2,(d_1,d_2)}|H_1 = h_1]$, and we compute $\hat{\sigma}_{i12,(d_1,d_2)} = \hat{E}^{(-k)}[\epsilon_{1,d_1}\epsilon_{2,(d_1,d_2)}|H_1 =$

$H_{i1}]$ for $i \in I^{(k)}_{(d_1,d_2)}$. The covariance matrix is then estimated as

$$\widehat{\Sigma}_i = \begin{pmatrix} \hat{\sigma}^2_{i1,d_1} & \hat{\sigma}_{i12,(d_1,d_2)} \\ \hat{\sigma}_{i12,(d_1,d_2)} & \hat{\sigma}^2_{i2,(d_1,d_2)} \end{pmatrix},$$

for each $i \in I^{(k)}_{(d_1,d_2)}$.

(d) Sample $\{(\tilde{\epsilon}_{i1}, \tilde{\epsilon}_{i2})\}_{i \in I^{(-k)}_{(d_1,d_2)}}$ independently from the empirical distribution of the standardized residuals $\{(\hat{\epsilon}_{i1,d_1}, \hat{\epsilon}_{i2,(d_1,d_2)})\widehat{\Sigma}_i^{-1/2}\}_{i \in I^{(-k)}_{(d_1,d_2)}}$, and compute $\epsilon_i^{\text{test}} = (\epsilon_{i1}^{\text{test}}, \epsilon_{i2}^{\text{test}})^\top = \widehat{\Sigma}_i^{-1/2}(\tilde{\epsilon}_{i1}, \tilde{\epsilon}_{i2})^\top$ for $i \in I^{(k)}_{(d_1,d_2)}$.

(e) Construct $(Q_{i1}^{\text{test}}, Q_{i2}^{\text{test}})^\top = (\hat{E}[Q_1(d_1)|H_1 = H_{i1}], \hat{E}[Q_2(d_1,d_2)|H_2 = H_{i2}])^\top + \epsilon_i^{\text{test}}$ for $i \in I^{(k)}_{(d_1,d_2)}$.

In this procedure, we estimate the conditional expectation functions of $Q_1(d_1)$ and $Q_2(d_1,d_2)$, as well as the conditional variances and covariance of the residuals $(\epsilon_{i1}, \epsilon_{i2})$, using random forests (Friedberg, Tibshirani, Athey, and Wager, 2021; Wager and Athey, 2018).

## A.8 Proof of Proposition 5.1

By the same reasoning as in the proof of Lemma 1 in Vuong and Xu (2017), under Assumption 5.1, equation (6) holds almost surely for any $d_1, d_1' \in \{U, T\}$. Moreover, under conditions (i) and (ii) of the random assignment of $(D_1, D_2)$, we have

$$Q_{S_2^k|(D_1,S_1)}(\cdot|d_1,S_1) = Q_{S_2^k(d_1)|S_1}(\cdot|S_1) \text{ and } F_{S_2^k|(D_1,S_1)}(\cdot|d_1,S_1) = F_{S_2^k(d_1)|S_1}(\cdot|S_1),$$

almost surely. Hence, $\widetilde{S}_2^k(d_1) = S_2^k(d_1)$ a.s.

Regarding the conditional average of the habit-formation effect in equation (5),

$$
\begin{aligned}
&E\left[\frac{\mathbf{1}\{D_1=T\}}{P\left(D_1=T|H_1\right)}\cdot\frac{\mathbf{1}\{D_2=U\}}{P\left(D_2=U|H_2\right)}\cdot Y_2\middle|\pi_1^*(H_1)=T,\pi_2^*(H_2)=U\right]\\
&=E\left[\frac{\mathbf{1}\{D_1=T\}}{P\left(D_1=T|H_1\right)}\cdot\frac{\mathbf{1}\{D_2=U\}}{P\left(D_2=U|H_2\right)}\cdot Y_2(T,U)\middle|\pi_1^*(H_1)=T,\pi_2^*(H_2(T))=U\right]\\
&=E\left[\frac{\mathbf{1}\{D_1=T\}}{P\left(D_1=T|H_1\right)}\cdot E\left[\frac{\mathbf{1}\{D_2=U\}}{P\left(D_2=U|H_2\right)}\middle|H_2\right]\cdot E\left[Y_2(T,U)|H_2\right]\middle|\pi_1^*(H_1)=T,\pi_2^*(H_2(T))=U\right]\\
&=E\left[\frac{\mathbf{1}\{D_1=T\}}{P\left(D_1=T|H_1\right)}\cdot Y_2(T,U)\middle|\pi_1^*(H_1)=T,\pi_2^*(H_2(T))=U\right]\\
&=E\left[Y_2(T,U)|\pi_1^*(H_1)=T,\pi_2^*(H_2(T))=U\right], \qquad (22)
\end{aligned}
$$

where the second line follows from condition (ii), and the last line follows from condition (i). It also follows that

$$
\begin{aligned}
&E\left[\frac{\mathbf{1}\{D_1=U\}}{P\left(D_1=U|H_1\right)}\cdot\frac{\mathbf{1}\{D_2=U\}}{P\left(D_2=U|H_2\right)}\cdot Y_2\middle|\pi_1^*(H_1)=T,\pi_2^*\left(\widetilde{H}_2(T)\right)=U\right]\\
&=E\left[\frac{\mathbf{1}\{D_1=U\}}{P\left(D_1=U|H_1\right)}\cdot\frac{\mathbf{1}\{D_2=T\}}{P\left(D_2=U|H_2\right)}\cdot Y_2(U,U)\middle|\pi_1^*(H_1)=T,\pi_2^*\left(H_2(T)\right)=U\right]\\
&=E\left[Y_2(U,U)|\pi_1^*(H_1)=T,\pi_2^*\left(H_2(T)\right)=U\right], \qquad (23)
\end{aligned}
$$

where the second line follows from the result $\widetilde{S}_2^k(d1)=S_2^k(d_1)$ a.s., and the last line follows by the same argument used in deriving (22). Combining these results yields (7). The results (8) and (7) can also be established by a similar argument. □

## A.9 Additional Tables

Table A.1: The external validity of the experimental sample

| | Experimental sample | Random sample of population | Difference between sample and population |
|---|---|---|---|
| Monthly electricity usage in July (kWh) | 362 | 304 | 58.33 |
| | [196] | [177] | (5.61) |
| Number of people at home | 2.51 | 2.31 | 0.20 |
| | [1.23] | [1.20] | (0.04) |
| Self-efficacy in energy conservation | 3.45 | 3.32 | 0.12 |
| (1-5 Likert scale) | [0.85] | [0.97] | (0.03) |
| Household income (JPY10,000) | 628 | 581 | 46.56 |
| | [380] | [384] | (11.45) |
| Number of households | 2,400 | 2,070 | |

Notes: We randomly sampled 2070 customers from the target population who did not participate in this experiment, and conducted a similar survey to the one for the experimental sample. The purpose of this survey was to investigate the external validity of our experimental sample by comparing the mean for each variable between our experimental sample and this random sample. Columns 1 and 2 show summary statistics for the experimental sample in summer and the random sample. Column 3 presents differences in means, with the standard errors of these differences in parentheses. We observe larger means for four variables in the experimental sample than in the random sample, and the differences are statistically significant. Our experimental sample has larger pre-experiment electricity usage per month, a larger number of people at home on weekdays, higher self-efficacy in energy conservation, and higher household income. This implies that our sample includes a larger number of customers who are willing and able to reduce their electricity consumption, which should be taken into consideration when discussing this study's generalizability. The monetary unit is given as 1 ¢ = 1 JPY in 2020.

Table A.2: Average Treatment Effects

| | $t = 1$ 2020 summer | $t = 2$ 2020 winter |
|---|---|---|
| Treated group ($D_t = T$) | -0.051 | -0.032 |
| | (0.013) | (0.013) |
| Number of customers | 2400 | 2400 |
| Number of observations | 591028 | 669212 |

Notes: This table shows the estimation results for equation 20. The dependent variable is the log of household-level electricity consumption over 30-minute intervals for summer 2020 in the first column and for winter 2020 in the second column. We include household fixed effects and time fixed effects for each 30-minute interval. The standard errors are clustered at the household level to adjust for serial correlation.

Table A.3: Dynamic Heterogeneity among Observables in the Second Period

| | | In pre-experiment, 2020 summer | | | | In pre-experiment, 2020 winter | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Peak hour usage - Pre-peak hour usage | | Peak hour usage - Post-peak hour usage | | Peak hour usage - Pre-peak hour usage | | Peak hour usage - Post-peak hour usage | |
| | All | Low | High | Low | High | Low | High | Low | High |
| $(D_1, D_2) = (U, T)$ | -0.046 | -0.069 | -0.029 | -0.027 | -0.070 | -0.084 | -0.020 | -0.047 | -0.045 |
| | (0.019) | (0.027) | (0.026) | (0.026) | (0.028) | (0.029) | (0.024) | (0.028) | (0.026) |
| $(D_1, D_2) = (T, U)$ | -0.011 | -0.006 | -0.017 | -0.022 | 0.002 | -0.001 | -0.016 | -0.022 | 0.000 |
| | (0.019) | (0.026) | (0.028) | (0.026) | (0.028) | (0.028) | (0.025) | (0.028) | (0.026) |
| $(D_1, D_2) = (T, T)$ | -0.029 | -0.036 | -0.027 | -0.044 | -0.013 | -0.033 | -0.022 | -0.015 | -0.042 |
| | (0.019) | (0.028) | (0.026) | (0.025) | (0.028) | (0.029) | (0.024) | (0.028) | (0.026) |
| Number of customers | 2400 | 1201 | 1199 | 1200 | 1200 | 1200 | 1200 | 1200 | 1200 |
| Number of observations | 669212 | 326214 | 342998 | 348751 | 320461 | 318280 | 350932 | 321540 | 347672 |
| p-value ($(T, U) = 0$) | 0.569 | 0.830 | 0.545 | 0.392 | 0.933 | 0.962 | 0.541 | 0.436 | 0.992 |
| p-value ($(U, T) = (T, T)$) | 0.351 | 0.245 | 0.930 | 0.496 | 0.037 | 0.062 | 0.928 | 0.238 | 0.905 |

| | Number of people at home (1 PM - 5 PM) | | Number of people at home (5 PM - 9 PM) | | Self-efficacy | | Household income | |
|---|---|---|---|---|---|---|---|---|
| | Low | High | Low | High | Low | High | Low | High |
| $(D_1, D_2) = (U, T)$ | -0.037 | -0.060 | -0.062 | -0.027 | -0.006 | -0.091 | -0.071 | -0.021 |
| | (0.025) | (0.029) | (0.026) | (0.027) | (0.026) | (0.027) | (0.028) | (0.026) |
| $(D_1, D_2) = (T, U)$ | -0.004 | -0.022 | -0.015 | 0.002 | -0.022 | 0.001 | -0.026 | 0.009 |
| | (0.025) | (0.029) | (0.026) | (0.028) | (0.027) | (0.027) | (0.027) | (0.028) |
| $(D_1, D_2) = (T, T)$ | -0.048 | -0.005 | -0.032 | -0.027 | -0.053 | -0.004 | -0.027 | -0.031 |
| | (0.025) | (0.029) | (0.026) | (0.027) | (0.026) | (0.027) | (0.027) | (0.027) |
| Number of customers | 1425 | 975 | 1370 | 1030 | 1236 | 1164 | 1264 | 1136 |
| Number of observations | 387296 | 281916 | 371509 | 297703 | 347489 | 321723 | 348331 | 320881 |
| p-value ($(T, U) = 0$) | 0.878 | 0.463 | 0.555 | 0.935 | 0.412 | 0.966 | 0.327 | 0.736 |
| p-value ($(U, T) = (T, T)$) | 0.664 | 0.042 | 0.232 | 0.990 | 0.064 | 0.001 | 0.087 | 0.716 |

Notes: This table shows the estimation results for equation 21 using the full-sample (the first column of the upper panel) or sub-samples (the remaining columns). The dependent variable is the log of household-level electricity consumption over 30-minute intervals in winter 2020. We include household fixed effects and time fixed effects for each 30-minute interval. The standard errors are clustered at the household level to adjust for serial correlation. To investigate the heterogeneity of the treatment effects, we focused on the variables selected for estimating the optimal policy in Section 4.2 and divided the sample into eight sets of two sub-groups. For the eight different variables, the first subgroup includes households who are below the median of this variable and the second includes those who are above the median. The monetary unit is given as 1 ¢=1 JPY in 2020.