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

We would like to thank Linnea Holy for her exceptional research assistance and Christopher Costello, Michael Greenstone, Ryan Kellogg, Anna Russo and seminar participants at the Coase Project Conference for their helpful comments. We thank the Japanese Ministry of Environment 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.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

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
JEL No. C1,C50,C9,C93,Q4,Q40,Q41,Q48,Q50,Q58

ABSTRACT

Economic policies often involve dynamic interventions, where individuals receive repeated interventions over multiple periods. This dynamics makes past responses informative to predict future responses and ultimate outcomes depend on the history of interventions. Despite these phenomena, existing economic studies typically focus on static targeting, possibly overlooking key information from dynamic interventions. We develop a framework for designing optimal dynamic targeting that maximizes social welfare gains from dynamic policy intervention. Our framework can be applied to experimental or quasi-experimental data with sequential randomization. We demonstrate that dynamic targeting can outperform static targeting through several key mechanisms: learning, habit formation, and screening effects. We then propose methods to empirically identify these effects. By applying this method to a randomized controlled trial on a residential energy rebate program, we show that dynamic targeting significantly outperforms conventional static targeting, leading to improved social welfare gains. We observe significant heterogeneity in the learning, habit formation, and screening effects, and illustrate how our approach leverages this heterogeneity to design optimal dynamic targeting.

Takanori Ida
Graduate School of Economics
Kyoto University
Yoshida, Sakyo
Kyoto 606-8501
Japan
ida@econ.kyoto-u.ac.jp

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

Koichiro Ito
Harris School of Public Policy
University of Chicago
1155 East 60th St
Chicago, IL 60637
and NBER
ito@uchicago.edu

Daido Kido
Otaru University of Commerce
3-5-21Midori
Otaru, Hokk 047-8501
Japan
dkido@res.otaru-uc.ac.jp

Toru Kitagawa
Department of Economics
Brown University
Box B
Providence, RI 02912
and Institute For Fiscal Studies - IFS
toru_kitagawa@brown.edu

Shosei Sakaguchi
Faculty of Economics
The University of Tokyo
7-3-1, Hongo
Bunkyo-ku
Tokyo 113-0033
Japan
sakaguchi@e.u-tokyo.ac.jp

Shusaku Sasaki
Osaka University
1-1 Yamadaoka
Suita, Osak 5650871
Japan
ssasaki.econ@gmail.com

1 Introduction

Targeting has become a central question in economics and policy design because policymakers who face budget constraints need to identify who should be treated to optimize policy impacts. The advent of machine learning and econometric methods has spurred research on targeting across various policy domains, including job training programs (Kitagawa and Tetenov, 2018), 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 Takana, 2023), and energy rebate programs (Ida, Ishihara, Ito, Kido, Kitagawa, Sakaguchi, and Sasaki, 2023).

While existing economic literature primarily focuses on *static* targeting, many real-world economic policies involve *dynamic* interventions. These policies, such as job training programs (Lechner, 2009; Rodríguez et al., 2022), unemployment insurance programs (Meyer, 1995; Kolsrud et al., 2018), healthcare programs (Luckett et al., 2019), and educational interventions (Ding and Lehrer, 2010), administer repeated interventions over multiple periods. The responses of individuals to these interventions can vary based on the timing and frequency of interventions. Moreover, data from earlier interventions can inform future responses. By leveraging these mechanisms and their heterogeneity, dynamic targeting could surpass static targeting. Thus, it is crucial to explore the design of optimal dynamic targeting and understand its potential superiority over static targeting, which may overlook key information.

This study demonstrates how researchers can utilize experimental or quasi-experimental data with sequential randomization to design dynamic targeting that maximizes social welfare gains from a dynamic policy intervention. We apply our framework to a randomized controlled trial (RCT) on an energy rebate program to empirically investigate the importance of dynamic targeting in enhancing welfare from dynamic policy intervention. Our framework draws insights from the dynamic treatment regimes literature in biostatistics (Robins, 1986; Murphy, 2003; Chakraborty and Moodie, 2013; Zhang et al., 2018; Tsiatis et al., 2019). In medical science, determining the optimal individualized timing and prescriptions of treatments to improve health outcomes is often crucial (Pelham Jr et al., 2016). This parallels our research question in social science: determining the optimal individualized timing and allocations of policy interventions to enhance social welfare gains.

Our dynamic targeting is a sequence of individualized treatment allocation policies that utilize two types

of information to optimize treatment assignment. The first type includes pre-intervention data such as household demographics, housing characteristics, and historical electricity usage. The second type encompasses the history of interventions for each individual and their responses to these interventions.

We propose several reasons why dynamic targeting could surpass static targeting, which we validate theoretically and empirically with experimental data. The first reason is the *learning effect*. Individuals who receive an intervention for multiple periods may learn how to respond effectively to the treatment (e.g., learning to conserve electricity in response to a rebate incentive). For those with a pronounced learning effect, repeated treatments could enhance social welfare gains compared to a single early-period treatment.

The second reason is the *habit formation effect*. Individuals may develop habits of conserving electricity when exposed to the rebate program. Consequently, even without participating in the program in the subsequent periods, they might consume less electricity than those who did not participate in an earlier period. For individuals with strong habit formation, early treatments could yield sustained welfare gains, even without subsequent treatments.

The third reason is the *screening effect*. Dynamic targeting allows us to gather new information about each individual after each intervention, which can inform treatment decisions in later periods. For instance, in a two-period binary choice experiment, we could assign an individual to treatment or no-treatment in the first period and observe their response. The response to treatment in the first period might be informative for optimizing their second-period assignment for some individuals. For others, observing the response to no-treatment in the first period could be more informative. In extreme cases, we might assign a first-stage treatment that yields statically sub-optimal welfare gain if the response to that treatment provides valuable information for optimizing assignments in subsequent periods.

Our theoretical framework elucidates the variation required by an RCT or quasi-experiment to estimate the optimal policy assignment. Following Sakaguchi (2021), we employ the Empirical Welfare Maximization (EWM) approach (Kitagawa and Tetenov (2018)) and Outcome Weighted Learning (Zhao et al. (2012)) to estimate dynamic targeting. We apply this method to data from an RCT for a residential electricity rebate program, which was conducted in collaboration with the Japanese Ministry of Environment. 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 can vary across individuals and can be positive, negative, or zero, considering the per-household implementation cost. This suggests that optimal targeting could enhance the social welfare

gain from this program.

We conducted 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. The data from these groups are instrumental in estimating dynamic targeting and examining its mechanism.

We empirically compare the welfare benefits of four non-targeting policies, two static targeting policies, and dynamic targeting in Section 4. Our findings indicate that conventional static targeting significantly enhances welfare benefits compared to non-targeting policies. Moreover, dynamic targeting can further augment these benefits, nearly doubling the welfare gain from static targeting policies.

We then explore the mechanism underpinning dynamic targeting in Section 5. We find that three potential mechanisms—learning, habit formation, and screening—significantly influence the optimal allocation of individuals to different policy interventions. Substantial heterogeneity exists among individuals for each of these effects. Dynamic targeting leverages this heterogeneity to construct optimal dynamic targeting, a strategy not employed by static targeting.

Related Literature and Our Contributions—Our study intersects with three areas of literature. First, numerous recent economics studies, including Johnson, Levine, and Toffel (forthcoming); Murakami, Shimada, Ushifusa, and Ida (2022); Cagala, Glogowsky, Rincke, and Strittmatter (2021); Christensen, Francisco, Myers, Shao, and Souza (2021); Assunção, McMillan, Murphy, and Souza-Rodrigues (2023); Gerarden and Yang (2023), have examined targeting in various policy domains. However, to our knowledge, existing economics studies primarily focus on static targeting. Our paper is among the first in economics to develop a framework for dynamic targeting in economic policy interventions. Sakaguchi (2021) studies estimation of dynamic treatment assignment rules under policy constraints, and as an application of the methods developed therein, he estimates the optimal allocation of students to classes with or without additional teacher aide across multiple grades. In the marketing literature, Ko et al. (2022) and Liu (2022) investigate dynamic coupon distributions in an online retail site using the Q-learning method.

Second, the medical statistics literature has explored the dynamic treatment assignment of medical interventions, known as the dynamic treatment regime. Our estimation of dynamic targeting policy builds on Sakaguchi (2021) with policy tree of Zhou et al. (2023).[1] Our paper is the first to offer theoretical and empir-

[1] Q-learning (Watkins and Dayan (1992); Murphy (2005)) is also a popular approach to estimate optimal dynamic treatment regimes. This approach can yield dynamic treatment regimes with low welfare when models relevant to outcomes are misspecified, even when experimental data is used.

ical comparisons between static and dynamic targeting and to investigate the mechanisms that could explain why dynamic targeting could outperform static targeting. Specifically, we derive a novel decomposition of the welfare gain into several key channels, clarifying the advantages of dynamic targeting, including the learning, habit formation, and screening effects. Additionally, we develop a novel approach to separately identify and estimate these effects using our experimental variation.

Third, this study contributes to the econometrics and statistics literature on dynamic treatment and mediation analysis. Heckman and Navarro (2007) and Heckman et al. (2016) explore the identification of dynamic treatment effects by modeling the selection process of treatments over multiple periods. Their focus on dynamic causal effects differs from ours as they do not consider the estimation of a dynamic targeting policy. Han (2023) examines the partial identification of the welfare impact of dynamic targeting in the presence of observational data subject to selection. 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 screening 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 et al. (2005)). We propose a novel identification approach for the screening 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.

Finally, the research question and policy environment examined in our study differ from those in the multi-armed bandit method. See Lattimore and Szepesvári (2020) for a recent monograph on the topic. See also Dimakopoulou et al. (2017), Ariu et al. (2021), Kasy and Sautmann (2021), and Kock et al. (2022) for recent developments of bandit algorithms with relevance to economics. The multi-armed bandit problem involves different individuals arriving at each period, each receiving treatment only once. In contrast, our study focuses on a policy environment where the same set of individuals arrive at each period, and each could receive multiple treatments across periods. Furthermore, in the multi-armed bandit problems, the treatment effect is explored and exploited across sequential periods, while in our framework, the effects of sequential treatments are estimated before the allocation task. Thus, our framework is distinct from the multi-armed bandit method.

2 Conceptual Framework

In this section, we introduce a theoretical framework for dynamic policy targeting. We formulate the problem in Section 2.1 and discuss the advantages of dynamic over static targeting in Section 2.3.

2.1 Dynamic Targeting

Consider a planner introducing a policy intervention to a heterogeneous population. The planner assigns a binary treatment to individuals, exploiting their diverse responses. Unlike the standard statistical treatment choice as in Manski (2004), Hirano and Porter (2009), and Kitagawa and Tetenov (2018), the planner can assign different treatments to the same individual over multiple periods. For clarity, we limit our analysis to two-period assignments, where the planner assigns treatments in two time periods, $t = 1$ and $t = 2$.[2]

2.1.1 Potential Outcomes

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

The planner aims to optimize a social welfare criterion by assigning each individual to one of the four intervention arms. 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. The outcomes represent individual welfare contributions, as defined in Section 4.1, rather than electricity consumption. As indicated by the arguments of these potential outcomes, we assume that the potential outcomes in period 1 are independent of the treatment d_2 in period 2. This assumption aligns with the no-anticipation condition, a common assumption in the literature of program evaluation with panel data. This condition stipulates that an individual's knowledge or expectation of the treatment to be given in period 2 does not influence their treatment response in period 1. This no-anticipation assumption also implies that an individual's knowledge or expectation of the treatment assignment rule in period 2 does not affect their treatment response in period 1. This guarantees the external validity of experimental data with sequentially randomized treatments for the population targeted by the optimal dynamic policy.

[2] Our analysis can be straightforwardly extended to more than two periods.

Let $Y_2(d_1, d_2)$ represent the potential outcomes in period 2 when an individual is assigned to intervention arm $(d_1, d_2) \in \{U, T\}^2$. By including period 1's treatment d_1 in the arguments of period 2's potential outcomes, we account for the dynamic causal effect of period 1's treatment on period 2's outcome. Empirical evidence, as shown below, supports the existence of this dynamic treatment effect in the context of an electricity rebate program, and its heterogeneity significantly influences the welfare performance of dynamic policy targeting.

2.1.2 Information available to the planner

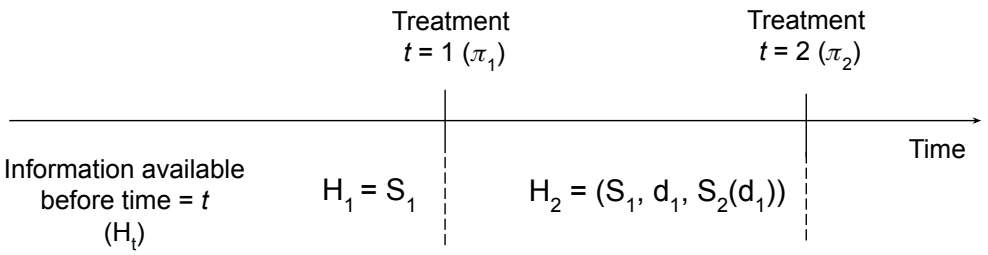
We introduce the sequential structure of the planner's information set and treatment assignments in dynamic policy targeting for an individual in the population.

At the beginning of period 1, the planner observes the individual's covariate information S_1 and assigns treatment $d_1 \in U, T$. In an electricity rebate program, S_1 includes past electricity usage, household demographics, and property characteristics. At the end of period 1, the planner observes the treatment response $Y_1(d_1)$, which is the post-treatment electricity usage.

At the beginning of period 2, the planner has updated covariate information S_2. Using (S_1, d_1) and S_2, the planner assigns the period 2 treatment . S_2 includes observable covariate information available after the first-period intervention and before the second-period intervention, including the first-period treatment response $Y_1(d_1)$. Since S_2 is influenced by period 1 treatment d_1, we use the potential outcome notation $S_2(d_1)$. In the electricity rebate program, $S_2(d_1)$ includes post-treatment electricity usage in period 1 and pre-treatment usage in period 2. Upon observing $S_2(d_1)$, the planner assigns period 2 treatment based on $(S_1, d_1, S_2(d_1))$.

In Figure 1, H_t denotes the information available to the planner before time t. At $t = 1$, H_1 is S_1, and at $t = 2$, $H_2(d_1) = (S_1, d_1, S_2(d_1))$. To consider counterfactual histories of period 1 treatment, we use the potential outcome notation for H_2.

Figure 1: Information Available to the Planner



Notes: This figure presents a set of information available to the planner before time t (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 $t = 1$.

Our framework accommodates full heterogeneity in treatment responses for periods 1 and 2. This heterogeneity makes a non-individualized uniform assignment suboptimal. We consider individualized sequential treatment assignments as follows: In period 1, the planner assigns $d_1 = T$ or U based on H_1. In period 2, upon observing $H_2(d_1)$, the planner assigns $d_2 = T$ or U. The targeting policy in period t is denoted as $\pi_t : \mathcal{H}_t \to T, U$, where $\mathcal{H}_t$ is the space of history in period t. A sequence of policies $\pi := (\pi_1, \pi_2)$ defines the sequential treatment assignment strategy. We call π a *dynamic targeting policy (DTP)*.

2.1.3 Optimal Dynamic Targeting

Given a fixed DT π, we specify the planner's social welfare criterion $W(\pi)$ of π as the population mean of the sum of individuals' outcomes over the two periods that are realized when treatment assignments follow π across the two stages:

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

Note that our social welfare sums up the flow outcomes over multiple periods. This differs from a setting in the dynamic treatment regime literature of medical intervention in which the health outcome of the terminal period matters for the welfare criterion (Robins, 1986; Murphy, 2003; Chakraborty and Moodie, 2013; Zhang et al., 2018; Tsiatis et al., 2019).

The optimal DT $\pi^* = (\pi_1^*, \pi_2^*)$ is obtained by:

$$\pi^* = \arg\max_{\pi\in\widetilde{\Pi}} W(\pi),$$

where $\widetilde{\Pi} = \widetilde{\Pi}_1 \times \widetilde{\Pi}_2$ with $\widetilde{\Pi}_t$ being a set of measurable functions π_t.

A DTP has two dynamic features in its sequential treatment assignment. First, it can switch intervention arms, U and T, between periods rather than assigning a unique intervention arm across all periods. For instance, DTPs can assign $(d_1, d_2) = (T, U)$ or (U, T) instead of being limited to (T, T) or (U, U). This flexibility can improve welfare compared to static policies that lock in a treatment over time.

Second, DTPs use updated information $H_2(d_1)$ to decide on interventions in the second period. The optimal DTP leverages the sequential arrival of new information to improve welfare, contrasting with static policies that use only the initial period's pre-intervention covariate information H_1. For example, if $Y_1(d_1)$ helps predict period 2 potential outcomes, an optimal DTP can incorporate it into the period 2 treatment assignment, unlike static targeting policies.

2.2 Static Targeting

As noted, we have two types of static targeting policies. The first type, *static targeting policy I* (STP-I), uses pre-intervention information H_1 to decide between the two fixed interventions (T, T) or (U, U). It is formulated as a map $\pi_{S(I)} : \mathcal{H}_1 \to (T, T), (U, U)$. The second type, *static targeting policy II* (STP-II), also uses H_1 but can choose among four interventions (U, U), (T, U), (U, T), or (T, T). It is formulated as a map $\pi_{S(II)} : \mathcal{H}_1 \to (U, U), (T, U), (U, T), (T, T)$.

For any STP, when the sequential intervention follows a fixed policy π_S, the average total outcomes are

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

This represents the social welfare of the static targeting policy π_S and is comparable to the social welfare of the DTP defined in (1). The optimal STP-I $\pi^*_{S(I)}$ and optimal STP-II $\pi^*_{S(II)}$ are the policies that maximize the social welfare function $W_S(\cdot)$ over all measurable STP-I and STP-II, respectively: $\pi^*_{S(j)} = \underset{\text{all measurable } \pi_{S(j)}}{\arg\max} W_S(\pi_{S(j)})$ for $j \in I, II$. The welfare attained by an optimal STP-II is no worse than that of an optimal STP-I (i.e., $W(\pi^*_{S(II)}) \geq W(\pi^*_{S(I)})$) because STP-I is limited to the set of policies satisfying $\pi_{S(I)} \in \{(T, T), (U, U)\}$.

The welfare attained by an optimal DTP π^* is no worse than that of an optimal STP-II (i.e., $W(\pi^*) \geq W_S(\pi^*_{S(II)})$) and that of the optimal STP-I because the optimal DTP utilizes the updated individual's covariate information $S_2(d_1)$, while STP-II does not. Hence, the welfare gain $W(\pi^*) - W_S(\pi^*_{S(II)})$ of the optimal

DTP relative to the optimal STP-II captures the welfare gain of having access to and using $(d_1, S_2(d_1))$ additionally in period 2 treatment choice.

2.3 Welfare Gains from Dynamic Targeting

This section analyzes the welfare gains of dynamic targeting relative to static targeting policies and interprets this in the context of the electricity conservation rebate program. We define channels of welfare gains from being dynamic: heterogeneity in dynamic causal effects of learning, habit formation, and the use of information to predict future treatment response. Section 5 empirically investigates the magnitude of each channel.

2.3.1 Learning

If individuals participate in a rebate program to save electricity, they acquire skills to reduce their electricity consumption. Consider individuals who participate in the rebate programs in the first and second periods. In the first period, they acquire the ability to save electricity. In the second period, they utilize this ability to further reduce their consumption. Consequently, those who participated in the first period saved more electricity in the second period than those who did not.

This effect is termed the *learning effect*, defined as $Y_2(T,T) - Y_2(U,T)$. We say an individual has a learning effect if $Y_2(T,T) > Y_2(U,T)$. Conversely, an individual exhibits a *habituation effect* if $Y_2(T,T) < Y_2(U,T)$. Individuals with a habituation effect become accustomed to the rebate program in the first period and thus do not perform as well in the second period.

For those with a learning effect, the sequential intervention (T,T) outperforms (U,T) in enhancing welfare contributions in the second period. The optimal DTP uses an individual's information H_1 and H_2 to identify those with a learning or habituation effect, whereas the optimal STI-II uses only first-period information H_1 for this purpose. The optimal STI-I cannot exploit the learning effect as it cannot assign (T,U).

2.3.2 Habit Formation and Other Effects

Some individuals may develop habits for saving electricity when exposed to the rebate program in the first period. For these individuals, the dynamic causal effect of period 1 treatment on the period 2 outcome

is positive, even if they are not exposed to the rebate program in the second period. This effect is termed the *habit formation effect*, defined as $Y_2(T,U) - Y_2(U,U)$. An individual is said to have a habit-formation effect if $Y_2(T,U) > Y_2(U,U)$.

Additionally, we introduce two alternative effects: the *second-period-treatment effect* ($Y_2(U,T) - Y_2(U,U)$) and the *full-intervention effect* ($Y_2(T,T) - Y_2(U,U)$). Combining these four effects (learning, habit formation, second-period-treatment, and full-intervention) enables us to characterize the (oracle) optimal sequential intervention for the second-period outcome as follows:

Remark 2.1. *An assignment $(d_1, d_2) \in \{U,T\}^2$ is optimal for the second period if and only if $Y_2(d_1,d_2) \geq Y_2(d_1', d_2')$ for any $(d_1', d_2') \in \{U,T\}^2$. The optimal assignments for the second period are exclusively characterized as follows:*

- *$(d_1,d_2) = (T,T)$ is optimal for the second period if and only if the learning effect ≥ 0, full-intervention effect ≥ 0, and full-intervention effect $\geq$ the habit-formation effect;*
- *$(d_1,d_2) = (T,U)$ is optimal for the second period if and only if the habit-formation effect ≥ 0, habit-formation effect $\geq$ the second-period-treatment effect, and habit-formation effect $\geq$ the full-intervention effect;*
- *$(d_1,d_2) = (U,T)$ is optimal for the second period if and only if the second-period-treatment effect ≥ 0, second-period-treatment effect $\geq$ the habit-formation effect, and learning effect ≤ 0 (i.e., there is habituation effect);*
- *$(d_1,d_2) = (U,U)$ is optimal for the second period if and only if the habit-formation effect ≤ 0, second-period-treatment effect ≤ 0, and full-intervention effect ≤ 0.*

2.3.3 Screening

One advantage of utilizing $S_2(d_1)$ in dynamic targeting is the ability to use the period 1 treatment response for treatment allocation in the second period. If the planner assigns $d_1 = T$ in the first period, they can observe $H_2(T)$; if $d_1 = U$, they observe $H_2(U)$. Note that $H_2(U)$ and $H_2(T)$ contain different variables, as $S_2(d_1)$ includes potential variables indexed by d_1. Thus, they provide different information for predicting individuals' unobserved types and treatment response behaviors in the second period. For

instance, if $H_2(T)$ is more informative for predicting the heterogeneity of treatment effects in the second-period outcome than $H_2(U)$, it is more beneficial to observe $H_2(T)$ by allocating $d_1 = T$, despite a potential loss in first-period welfare, than to observe $H_2(U)$ by allocating $d_1 = U$. This channel of welfare gain is termed the *screening effect* of period 1 treatment. Given an optimal DTP π^*, the *screening effect* is defined as

$$\tau_{scr} \equiv Y_2(T, \pi_2^*(H_2(T))) - Y_2(T, \pi_2^*(H_2(U))). \tag{3}$$

If the first-period intervention is T and the second-period treatment follows the optimal policy π_2^*, the screening effect τ_{scr} represents the welfare gain from acquiring the information $H_2(T)$ instead of $H_2(U)$. A positive τ_{scr} indicates that $H_2(T)$ is more beneficial than $H_2(U)$ for selecting the optimal intervention in the second period.

Dynamic targeting can incorporate screening effects into sequential treatment choices. Note that screening effects cannot be identified even with RCT data of sequentially randomized treatment assignments, as we cannot observe the counterfactual history $H_2(U)$ in equation (3) for those treated in the first period. In Section 2.3.3, we impose a conditional rank-invariance assumption (Heckman et al., 1997; Chernozhukov and Hansen, 2005; Vuong and Xu, 2017) between $S_2(U)$ and $S_2(T)$ and present identification and estimation of the average screening effects.

2.3.4 Decomposition of Welfare Gains

This section presents a decomposition formula of the welfare gain of a DTP relative to a status-quo if uniform no-treatment, $W(\pi) - W_{(U,U)}$, where $W_{(U,U)} \equiv E[Y_1(U) + Y_2(U, U)]$.

The next proposition shows how the welfare gain of the optimal DTP π^* is decomposed into the sources of welfare gain listed in the previous sections. The Appendix presents the proof.

Proposition 2.1. *The welfare gain of an optimal DTP relative to no-intervention status-quo admits the*

following decomposition:

$$
\begin{aligned}
& W(\pi^*) - W_{(U,U)} \\
& = \underbrace{E[Y_1(T) - Y_1(U) | \pi_1^*(H_1) = T]}_{\textit{Treatment effect on the treated in } t=1} \cdot \Pr(\pi_1^*(H_1) = T) \\
& + \underbrace{E[Y_2(U,T) - Y_2(U,U) | \pi_2^*(H_2(U)) = T]}_{\textit{Treatment effect on the treated in } t=2} \cdot \Pr(\pi_2^*(H_2(U)) = T) \\
& + \underbrace{E[Y_2(T,U) - Y_2(U,U) | \pi_1^*(H_1) = T, \pi_2^*(H_2(T)) = U]}_{\textit{Habit formation effect for those assigned to } (T,U)} \cdot \Pr(\pi_1^*(H_1) = T, \pi_2^*(H_2(T)) = U) \\
& + \underbrace{E[Y_2(T,T) - Y_2(U,T) | \pi_1^*(H_1) = T, \pi_2^*(H_2(T)) = T]}_{\textit{Learning effect for those assigned to } (T,T)} \cdot \Pr(\pi_1^*(H_1) = T, \pi_2^*(H_2(T)) = T) \\
& + \underbrace{E[Y_2(T, \pi_2^*(H_2(T))) - Y_2(T, \pi_2^*(H_2(U))) | \pi_1^*(H_1) = T]}_{\textit{Screening effect on the treated in } t=1} \cdot \Pr(\pi_1^*(H_1) = T). \qquad (4)
\end{aligned}
$$

The results indicate that the welfare gain from the optimal DTP π^* can be broken down into five distinct effects. These include the average treatment effect for those treated in period 1, the average treatment effect for those treated in period 2, the average habit-formation effect for individuals assigned to (T, U), the average learning effect for individuals assigned to (T, T), and the average screening effect for those treated in the first period across several groups. These groups are characterized by the optimal policy π^* and are weighted according to the proportions of the corresponding conditioning groups. This decomposition is, to the best of our knowledge, a novel contribution to the literature and encapsulates all sources of welfare gains from DTP. As highlighted in the proof of Proposition 2.1, this decomposition does not necessitate the welfare of DTP to be optimal. The same decomposition can be achieved for any arbitrary DTP π. In Sections 4 and 5, we examine the identification of each welfare gain factor, utilizing RCT data with sequentially randomized treatments.

3 Field Experiment and Data

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

3.1 Field Experiment

In 2020, we conducted a two-period field experiment in the Kansai and Chubu regions of Japan, in collaboration with the country's Ministry of the Environment.[3] We recruited a diverse group of households via letter and e-mail, offering a participation fee of 2000 JPY (≈20 USD; 1¢≈1 JPY in summer 2020). Of the 3470 customers who pre-registered, we excluded nonresidential customers, those who terminated their electricity contracts during the experiment, and those with incomplete high-frequency electricity usage data. This resulted in a final sample of 2400 residential customers. Our experiments were RCTs for consenting households, a common approach in residential electricity demand literature (Wolak, 2006, 2011; Ito, Ida, and Tanaka, 2018).

During the summer and winter of 2020, we randomly assigned these 2400 households to either the untreated group (U) or the treated group (T).[4]

Untreated group (U): Customers did not participate in the rebate program.

Treated group (T): Customers participated in the rebate program.

After two rounds of randomization, customers fell into one of four groups: (U, U), assigned to be untreated in the first (summer in 2020) and second (winter in 2020) periods; (U, T), assigned to be treated in the first period and treated in the second period; (T, U), assigned to be treated in the first period and untreated in the second period; and (T, T), assigned to be treated in the first and second periods. The number of households is 625 for (U, U), 606 for (U, T), 581 for (T, U), and 588 for (T, T).

Our experiment utilized a "peak-time rebate" (PTR) program (Wolak, 2011; Ida, Ishihara, Ito, Kido, Kitagawa, Sakaguchi, and Sasaki, 2023). 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.

Even so, PTR is generally considered less effective than dynamic pricing, where hourly prices equal c for two reasons. First, while PTR incentivizes customers to reduce electricity usage, it does not "penalize" customers for increasing their usage. Second, if policymakers do not carefully set the "baseline usage," PTR may inadvertently incentivize customers to manipulate their baseline usage to earn a larger rebate. Hence,

[3] 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).

[4] The random assignment process was designed such that $U : T = 1 : 1$ for each of the two periods.

economists often favor dynamic pricing over PTR (Wolak, 2011; Ito, Ida, and Takana, 2023).

However, political feasibility often hinders the implementation of dynamic pricing across the general population. The PTR is politically favored as it does not financially burden customers. This preference led the Japanese government, our experimental partner, to focus on PTR rather than dynamic pricing in this research project. They implemented a similar policy in the real world in the summer of 2022.[5]

The goal of our PTR was to decrease residential electricity consumption during system peak hours (1 pm to 5 pm in the summer; 5 pm to 9 pm in the winter) during the weeks of August 24–30, 2020 (period 1) and December 14–20, 2020 (period 2). During these treatment weeks, customers in the rebate program received a rebate equivalent to their energy conservation during peak hours relative to their baseline usage (in kWh), multiplied by 100 JPY per kWh. The baseline usage was determined by each customer's average electricity usage during peak hours in July or November. Customers were informed of the treatment week, peak hours, and reward calculation procedure at the start of August or December, preventing them from manipulating their baseline usage.

3.2 Data and Summary Statistics

Data were collected 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.[6] Columns 1–4 display the sample averages for each random assignment group in each period, with standard deviations in brackets. Column 5 presents the p-values of the F-test for differences in sample averages across the four groups.

[Table 1 about here]

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

[5] In June 2022, the Japanese government launched an electricity rebate program called the "Setsuden Point Program" to address an anticipated electricity supply shortage. This program, like the one in our field experiment, offered rebates to customers who reduced their electricity usage. California employed similar electricity rebate programs during the electricity crisis of 2000–2002 and in subsequent years (Reiss and White, 2008; Ito, 2015).

[6] Appendix A.2 presents empirical evidence of the average treatment effects and the dynamic heterogeneity in the treatment effects on peak-hour electricity usage, suggesting that dynamic policy targeting, as discussed in Section 4, could significantly enhance the policy outcome.

midnight). The remaining variables, derived from the surveys, 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 estimates the optimal DTP using our experimental data from the energy-saving rebate program. Section 4.1 details the construction of a social welfare criterion in our empirical context. Section 4.2 explains the empirical estimation of the optimal static and dynamic targeting policies using our experimental data. The estimation results are reported in Section 4.3.

4.1 Construction of the Social Welfare Criterion

We here define the explanatory variables for our estimation. We denote the price and marginal cost of electricity as p and c, respectively. In peak hours, the time-invariant residential price p is often lower than the marginal cost c. The aim of PTR programs is to mitigate the welfare loss resulting from this economic inefficiency by aligning the rebate incentive with c.

Consider a household that joins the rebate program. For each sequential intervention $(d_1, d_2) \in \{T, U\}^2$, we denote the potential consumption of electricity in periods 1 and 2 as $Q_1(d_1)$ and $Q_2(d_1, d_2)$, respectively. We assume a locally-linear demand curve for electricity usage. Then, the short-run welfare contributions of the intervention (d_1, d_2) in periods 1 and 2 can be, respectively, written by $\frac{1}{2}(p-c)(Q_1(d_1) - Q_1(U))$ and $\frac{1}{2}(p-c)(Q_2(d_1, d_2) - Q_2(U, U))$. Further, we consider that the reduction in consumption creates an additional long-run welfare contribution as it saves the cost of power plant investments. We denote this long-run gains of the intervention (d_1, d_2) in periods 1 and 2 by $\delta(Q_1(d_1) - Q_1(U))$ and $\delta(Q_2(d_1, d_2) - Q_2(U, U))$, respectively, where δ is the price per kW in the capacity market. Finally, participation in the rebate program in each period incurs an implementation cost per customer by a.

Thus, the welfare contribution of the intervention (d_1, d_2) from the rebate program for periods 1 and 2

can be, respectively, written by

$$\Delta Y_1(d_1) \equiv b \cdot (Q_1(d_1) - Q_1(U)) - a \cdot 1\{d_1 = T\}, \tag{5}$$

$$\Delta Y_2(d_1, d_2) \equiv b \cdot (Q_2(d_1, d_2) - Q_2(U, U)) - a \cdot 1\{d_2 = T\}, \tag{6}$$

with $b = \frac{1}{2}(p - c) + \delta$. $\Delta Y_1(d_1)$ and $\Delta Y_2(d_1, d_2)$ indicate how much an individual contributes to the social welfare when the individual receives the intervention (d_1, d_2) from the program. No-treatment $d_1 = U$ in period 1 incurs zero welfare contribution in period 1 (i.e., $\Delta Y_1(U) = 0$), and no-treatment $(d_1, d_2) = (U, U)$ across the two periods incurs zero welfare contribution in period 2 (i.e., $\Delta Y_2(U, U) = 0$).

We define *welfare gains* of a DTP π and STP π_S as equations (1) and (2) with $Y_1(d_1)$ and $Y_2(d_1, d_2)$ replaced by $\Delta Y_1(d_1)$ and $\Delta Y_2(d_1, d_2)$, respectively:

$$\Delta W(\pi) \equiv E\left[\sum_{(d_1,d_2)\in\{T,U\}^2} (\Delta Y_1(d_1) + \Delta Y_2(d_1, d_2)) \cdot 1\{\pi_1(H_1) = d_1, \pi_2(H_2(d_1)) = d_2\}\right]; \tag{7}$$

$$\Delta W_S(\pi_S) \equiv E\left[\sum_{(d_1,d_2)\in\{T,U\}^2} (\Delta Y_1(d_1) + \Delta Y_2(d_1, d_2)) \cdot 1\{\pi_S(H_1) = (d_1, d_2)\}\right]. \tag{8}$$

The welfare gains represent how much the targeting policies improve social welfare. Note that $Q_1(U)$ and $Q_2(U, U)$ in equations (5) and (6) do not depend on policy assignment. Therefore, the DTP and STP that maximize the welfare gains—$\Delta W(\cdot)$, $\Delta W_S(\cdot)$—are obtained as the DTP and STP that maximize the welfare functions (1) and (2) with $Y_1(d_1) \equiv b \cdot Q_1(d_1) - a \cdot 1\{d_1 = T\}$ and $Y_2(d_1, d_2) \equiv b \cdot Q_2(d_1, d_2) - a \cdot 1\{d_2 = T\}$.

Our social welfare criterion comprises four exogenous parameters: p, c, a, and δ. We use data from the Japanese electricity market during our experimental period to set the values for these parameters. p is the unit price of electricity. We set the regulated price of electricity in Japan, which is independent of the time of day, to $p = 25$ JPY/kWh.[7] c is the marginal cost of production for electricity. We specify $c = 125$ JPY/kWh such that the difference between p and c is equal to the rebate per kWh, which is 100 JPY. The wholesale price of electricity sometimes soars during peak hours, such as summer afternoons or

[7] In Japan, until April 1, 2016, household electricity was supplied by local power companies, and retail prices were regulated. Since then, entry into the retail electricity industry has been fully liberalized, allowing all households to freely choose their price menu. However, as a transitional measure, the regulated price for households is being maintained for now and set at approximately 25 JPY/kWh regardless of the time of day.

winter evenings, reflecting supply constraints. In the past, the wholesale price has occasionally exceeded 100 JPY/kWh in summer afternoons.[8]

Parameter a represents the administrative cost of implementing our energy-saving program. This cost comprises several items, including the installation cost of the Home Energy Management System (HEMS) required to participate. In 2016, the Japanese government estimated the cost of implementing a demand reduction program, including the installation cost of HEMS, to be 291.1 JPY per household per season (Ida and Ushifusa, 2017).[9] Because we consider the problem of two periods, we use half of this (291.1/2 JPY) as the value of the administrative cost per period.

Parameter δ represents the long-term benefits of a unit reduction in energy consumption. Here, we consider the effect of a unit reduction on the capacity market, where future supply capacity is traded between the power generation and retail sectors. In Japan, the capacity market was established in 2020, with the first auction held at that time. In that auction, the Japanese government provided a reference price of 9425 JPY/kW to bidders, which we use as the value for δ.

4.2 Estimation: Dynamic and Static Empirical Welfare Maximization

We estimate the optimal DTP and SPTs using our experimental data $\{(D_{i1}, D_{i2}, Y_{i1}, Y_{i2}, S_{i1}, S_{i2})\}_{i=1}^{n}$, with the sample size $n = 2400$. The outcome Y_{it} represents the observed welfare contribution and is constructed as $Y_{it} = b \cdot Q_{it} - a \cdot 1\{D_{it} = T\}$, where Q_{it} denotes the peak-time electricity consumption in the event period in period t for household i. Regarding the pre-intervention information S_{i1} $(= H_{i1})$, we use five variables in S_{i1}: Peak-time baseline electricity consumption in summer, household income, the number of household members usually at home from 13:00 to 17:00 on weekdays, the number of household members usually at home from 17:00 to 21:00 on weekdays, and a measure of the households' self-efficacy in energy conservation. Regarding the updated information S_{i2} in period 2, we use the peak-time electricity consumption in the event period in summer (Q_{i1}) and peak-time baseline electricity consumption in the winter. These variables are selected based on their ability to predict electricity consumption and the conditional

[8]The wholesale electricity market, where the power generation and retail sectors trade electricity, is operated by the Japan Electric Power Exchange (JEPX). Most trading occurs in the "day-ahead market" where both sectors trade electricity on the day before the actual demand period. Trading results are disclosed, and we confirm that the price exceeded 100 JPY/kWh on July 25, 2018. Moreover, the price has even exceeded 125 JPY/kWh. For example, the price reached 250 JPY/kWh on January 15, 2021.

[9]We do not include the installation cost for a smart meter in the administrative cost. Since the Great East Japan Earthquake of March 11, 2011, and the accident at the Fukushima Daiichi Nuclear Power Plant, the Japanese government has stipulated that smart meters should be installed in all homes by the end of the decade. Thus, this cost is "sunk" in that it will be paid regardless of whether a demand reduction program is implemented.

average treatment effects.

To estimate optimal DTP and SPTs-I and II, we employ Static and Dynamic Empirical Welfare Maximization (Dynamic and Static EWM) methods in Kitagawa and Tetenov (2018) and Sakaguchi (2021). First, consider estimating an optimal STP-I $\pi^*_{S(I)}$. Let $\Pi_{S(I)}$ be a pre-specified class of STPs-I $\pi_{S(I)}$ (e.g., class of decision trees). Using the RCT data, the EWM method estimates the optimal STP-I $\pi^*_{S(I)}$ by maximizing the empirical analog of the social welfare function (2) over $\Pi_{S(I)}$:

$$\begin{aligned} \hat{\pi}^*_{S(I)} &\in \underset{\pi_S \in \Pi_{S(I)}}{\arg\max}\, \widehat{W}_S(\pi_S), \\ \widehat{W}_S(\pi_S) &\equiv \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(Y_{i1}+Y_{i2})\cdot 1\{\pi_{S(I)}(H_{i1})=(D_{i1},D_{i2})\}}{P((D_1,D_2)=(D_{i1},D_{i2})|H_1=H_{i1})}\right), \end{aligned} \tag{9}$$

where $\widehat{W}(\pi_S)$ is an empirical welfare function of π_S that produces an unbiased estimate of the population social welfare $W_S(\pi_S)$. Observations are weighted by the inverse of the propensity scores, $P((D_1,D_2)=(D_{i1},D_{i2})|H_1=H_{i1})$, known from the RCT design. Regarding STPs-II, letting $\Pi_{S(II)}$ be a pre-specified class of STPs-II $\pi_{S(I)}$ (e.g., class of decision trees), we can apply the EWM to estimate the optimal STP-II $\pi^*_{S(II)}$ by replacing $\Pi_{S(I)}$ with $\Pi_{S(II)}$ in the EWM problem (9).

The EWM approach is model-free: It does not require any assumptions or a functional form specification for the potential outcome distributions. However, the policy class $\Pi_{S(I)}$ ($\Pi_{S(II)}$) must be specified. If the class $\Pi_{S(I)}$ ($\Pi_{S(II)}$) is too rich, the EWM solution $\hat{\pi}^*_{S(I)}$ ($\hat{\pi}^*_{S(II)}$) overfits the RCT data, and the social welfare attained by the estimated policy falls. We use decision tree classes (Breiman et al., 2017) for $\Pi_{S(I)}$ and $\Pi_{S(II)}$ because of the ease of interpretation of the decision tree-based assignment policies and the availability of partition search algorithms from the classification tree literature.

We proceed to elucidate the estimation of the optimal DTP π^*. We apply dynamic EWM, an extension of the EWM in Kitagawa and Tetenov (2018) to dynamic settings, as developed in Sakaguchi (2021). This method, implemented with backward induction, offers computational efficiency.[10] Let $\Pi \equiv \Pi_1 \times \Pi_2$ denote a pre-specified class of DTPs where Π_1 and Π_2 denote classes of policies in periods 1 and 2, respectively. The dynamic EWM with backward induction is a stepwise procedure for estimating the optimal DTP. It first estimates the optimal policy in the second period π^*_2 given each fixed $d_1 \in \{U,T\}$ by solving the EWM

[10]Two approaches exist for estimating an optimal DTP. The first, based on backward induction, estimates the optimal policy in the second period before proceeding to the first. The second approach, based on simultaneous maximization, estimates the entire DTP by maximizing the empirical analog of the welfare function of π. The backward induction approach we use offers computational advantages over the simultaneous maximization approach.

problem in the second period:

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

where $\widehat{W}_2(\pi_2)$ is an second-period empirical welfare function of π_2. Note that the weighting propensity score $P(D_2 = D_{i2}|H_2 = H_{i2})$ is known from the RCT design. The solution $\hat{\pi}_2^*$ is the estimate of the optimal policy in the second period π_2^*.

Subsequently, the method estimates the optimal policy in the first period π_1^* by solving the following EWM problem:

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

$\widehat{W}(\pi_1)$ is an estimator of the welfare function of π_1 when the optimal policy π_2^* is followed in the second period. Hence, the solution $\hat{\pi}_1$ estimates the optimal policy in the first period. As in the static EWM, we must specify the classes of period-specific policies Π_1 and Π_2 for estimation. We use a class of decision trees for each of Π_1 and Π_2. Through the two stepwise procedures, we obtain the DTP $\hat{\pi}^* = (\hat{\pi}_1^*, \hat{\pi}_2^*)$, the estimator of the optimal DTP π^*. Sakaguchi (2021) shows that the resulting DTP $\hat{\pi}^*$ is a consistent and rate-optimal estimator of the optimal DTP π^* when the class of DTPs Π contains the optimal DTP π^*.

For the dynamic targeting, we specify each of the period-specific policy classes, Π_1 and Π_2, to be the class of decision trees of depth 4. We also use the class of decision trees of depth 4 to each of the static-policy classes, $\Pi_{S(I)}$ and $\Pi_{S(II)}$, where $\Pi_{S(I)}$ is a class of policies with the two arms, (U, U) and (T, T), but $\Pi_{S(II)}$ is a class of policies with the four arms: (U, U), (T, U), (U, T), and (T, T). In the static EWM and each step of the dynamic EWM, we maximize the empirical welfare criterion exactly using a class of decision trees of depth 4, applying the exhaustive search algorithm of Zhou, Athey, and Wager (2023).[11]

[11] Given computational constraints, obtaining a globally optimal tree of depth 4 that exactly maximizes the empirical welfare is challenging. To mitigate this, we use a heuristic two-step procedure to approximate the globally optimal depth-4 tree. We first optimize a parent tree of depth 2 that maximizes the empirical welfare in the entire sample, dividing the sample into four subsamples. For each subsample, we search for a child depth-2 tree that maximizes the empirical welfare within the subsample. We then graft the child depth-2 trees onto the parent tree to construct the depth-4 tree. This grafted-tree approach is common in

Once the optimal DTP and STPs are estimated, we subsequently estimate the optimal welfare gains, $\Delta W(\pi^*)$, $\Delta W_S(\pi^*_{S(I)})$, and $\Delta W_S(\pi^*_{S(II)})$. One caveat of the EWM estimation is that the optimized empirical welfare value from the estimation will be an upwardly biased estimate of the true welfare attained by the estimated policy. This is known as the winner's bias (see, e.g., Andrews, Kitagawa, and McCloskey, 2019) and is caused by using the same data twice: once to learn the policy and once to infer the policy's welfare.[12] To control for the winner's bias in the welfare gain estimation, we create artificial test data by fitting a causal forest (Wager and Athey, 2018) to run regressions of the outcome onto all the covariates and generate data with permuted regression residuals. Appendix A.3 presents detailed explanations of our estimation procedure.

We denote the constructed test data by $\mathcal{S}^{\text{test}} \equiv \{S^{\text{test}}_{i1}, S^{\text{test}}_{i2}, Y^{\text{test}}_{i1}, Y^{\text{test}}_{i2}, D^{\text{test}}_{i1}, D^{\text{test}}_{i2} : i = 1, \ldots, n\}$, where $(D^{\text{test}}_{i1}, D^{\text{test}}_{i2}, S^{\text{test}}_{i1}, S^{\text{test}}_{i2})$ is identical to $(D_{i1}, D_{i2}, S_{i1}, S_{i2})$ in the original data. Note that the welfare gain of π can be decomposed as $\Delta W(\pi) = W(\pi) - W_U$, where $W_U \equiv E[Y_1(U) + Y_2(U,U)]$. With the test data, we obtain a point estimator for the welfare gain of the optimal DTP π^* and STPs, $\pi^*_{S(I)}$ and $\pi^*_{S(II)}$, as follows: For $\hat{\pi} = \hat{\pi}^*, \hat{\pi}^*_{S(I)}$, and $\hat{\pi}^*_{S(II)}$,

$$\widehat{\Delta W}(\hat{\pi}) \equiv \frac{1}{n}\sum_{i=1}^{n} \frac{1\{D^{\text{test}}_{i1} = \hat{\pi}_1(H^{\text{test}}_{i1})\}}{P(D_1 = D^{\text{test}}_{i1} | H_1 = H^{\text{test}}_{i1})} \left(Y^{\text{test}}_{i1} + \frac{1\{D^{\text{test}}_{i2} = \hat{\pi}_2(H^{\text{test}}_{i2})\}}{P(D_2 = D^{\text{test}}_{i2} | H_2 = H^{\text{test}}_{i2})} \cdot Y^{\text{test}}_{i2} \right)$$
$$- \frac{1}{n}\sum_{i=1}^{n} \frac{1\{D^{\text{test}}_{i1} = U\}}{P(D_1 = U | H_1 = H^{\text{test}}_{i1})} \left(Y^{\text{test}}_{i1} + \frac{1\{D^{\text{test}}_{i2} = U\}}{P(D_2 = U | H_2 = H^{\text{test}}_{i2})} \cdot Y^{\text{test}}_{i2} \right).$$

4.3 Results of the Optimal Policy Assignment

We estimate the welfare gains of the optimal DTP and STPs and non-targeting policies. We compare five alternative policies: 1) assigning everyone to (U,U), 2) assigning everyone to (T,U), 3) assigning everyone to (U,T), 4) assigning everyone to (T,T), 5) optimal STP-I $\pi^*_{S(I)}$, assigning each consumer to either (U,U) or (T,T) depending on the pre-intervention information H_1, 6) optimal STP-II $\pi^*_{S(II)}$, assigning each consumer to each of the four arms $(d_1, d_2) \in \{U,T\}^2$ depending on the pre-intervention information H_1, and 7) the optimal dynamic targeting π^*, adaptively assigning each consumer to each of the four arms depending on the updating information H_1 and H_2.

machine learning literature when constructing tree classifiers for computational feasibility. See, e.g., Chapter 2 of Breiman et al. (2017) and Section 9.2 in Hastie et al. (2009).

[12] The estimation and inference procedures proposed by Andrews, Kitagawa, and McCloskey (2019) cannot be directly applied to decision-tree-based policies because the number of candidate policies is infinite.

[Table 2 about here]

Table 2 presents the welfare performances of four benchmark policies without targeting (100% (U, U), 100% (T, U), 100% (U, T), and 100% (T, T)), followed by the optimal static targeting ($\pi^*_{S(I)}$ and $\pi^*_{S(II)}$) and dynamic targeting (π^*). For each policy, we estimate the welfare gain in JPY per household over the two periods. Among the four non-targeting policies, 100% (U, T) induces the highest welfare gain at 470.8 JPY per consumer, exceeding that of 100% (T, T). This would be because our policy intervention incurs costs (from implementation) and benefits (from energy conservation), making it suboptimal to always assign the rebate intervention.

The fact that the net welfare gain from a consumer can be positive, negative, or zero implies that policy performance could be increased through targeting. Table 2 shows that each of the three targeting policies achieves a higher welfare gain than the non-targeting policies. Comparing the two optimal STPs, STP II yields a higher welfare gain than STP I, implying that allowing for switching intervention between the two periods (i.e., allowing (T, U) and (U, T)) improves the social welfare (845.3 JPY versus 770.6 JPY). Further, Table 2 shows that dynamic targeting π^* achieves the highest social welfare (1684.3 JPY).

[Table 3 about here]

The DTP demonstrates a significantly higher welfare gain then each of the non-targeting policies and static targeting policies. Notably, the DTP yields a 839.1 JPY higher welfare gain than the optimal static targeting policy, $\pi^*_{S(II)}$. This difference is statistically significant. This result suggests that the adaptive treatment assignment depending on the updated information mostly improves the social welfare.

[Table 4 about here]

Noting that the welfare gain of dynamic targeting arises from two periods, Table 4 presents the decomposition of the results from Table 3 into periods 1 and 2. For the optimal DTP, the results reveal that 298.4 JPY (17.6%) of its welfare gain originates from period 1, whereas 1395.0 JPY (82.4%) of the welfare gain comes from period 2. Period 2 accounts for most of the welfare gain from dynamic targeting. Table 4 also indicates that the differences in welfare between dynamic targeting and static targeting primarily result from the second period.

We next investigate whether the estimated DTP $\hat{\pi}^*$ successfully assigns consumers to their optimal intervention groups. One way to do this is to estimate the counterfactual welfare difference of each group (d_1, d_2)

assigned by the estimated DTP $\hat{\pi}^*$ relative to each non-targeting policy. Specifically, for a group (d_1, d_2) assigned by the DTP $\hat{\pi}^*$ and a counterfactual intervention (d_1', d_2'), its counterfactual welfare difference is defined as

$$WD^{\hat{\pi}^*}_{(d_1,d_2),(d_1',d_2')} \equiv E[(Y_1(d_1) + Y_2(d_1, d_2)) - \big(Y_1(d_1') + Y_2(d_1', d_2')\big) \,|\hat{\pi}_1^*(H_1) = d_1, \hat{\pi}_1^*(H_2(d_1)) = d_2].$$

This is the welfare difference for those assigned to (d_1, d_2) by the DTP $\hat{\pi}^*$ if they deviate from the assigned group to the other group (d_1', d_2'). If the DTP $\hat{\pi}^*$ is (close to) optimal, for each assigned (d_1, d_2), the welfare difference $WD^{\hat{\pi}^*}_{(d_1,d_2),(d_1',d_2')}$ must be non-negative for any counterfactual intervention (d_1', d_2').

[Table 5 about here]

Table 5 shows estimation results for the counterfactual welfare differences $WD^{\hat{\pi}^*}_{(d_1,d_2),(d_1',d_2')}$ for each assigned group (d_1, d_2) and each counterfactual intervention (d_1', d_2'). Each counterfactual welfare difference is estimated to be positive except for that of $(d_1, d_2) = (U, U)$ and $(d_1', d_2') = (T, U)$. Some welfare differences are statistically and significantly positive. For example, the individuals belonging to the optimal assignment group (T, T) have a statistically significant welfare gain of 1819.7 JPY relative to a hypothetical assignment to (U, U). These results imply that the estimated DTP $\hat{\pi}^*$ successfully assigns consumers to the optimal intervention groups on average for the most part. The only anomaly is when the individuals assigned to (U, U) by the estimated DTP are hypothetically assigned to (T, U), but this welfare difference $WD^{\hat{\pi}^*}_{(U,U),(T,U)}$ (-271.7 JPY) is not statistically significant.

5 Mechanism Behind the Dynamic Targeting

This section investigates the mechanism of welfare improvement by the optimal DTP π^* using the decomposition outlined in equation (4). By examining each term of this decomposition, we can ascertain the contribution of each of the five effects to the welfare gain of the optimal DTP π^*. These effects are the first-period treatment effect, the second-period treatment effect, the habit-formation effect, the learning effect, and the screening effect. However, the conditional averages of the habit-formation effect, learning effect, and screening effect in equation (4) are counterfactual. They cannot be identified without an additional assumption. Section 5.1 proposes a novel approach to identifying and estimating these components of the

decomposition in equation (4). Finally, Section 5.2 presents the estimation results for the decomposition in equation (4) and investigates the mechanism behind the optimal DTP π^*.

5.1 Identification

We aim to identify and estimate each term comprising equation (4). However, the following components in equation (4) are counterfactual and cannot be identified without relying on an additional assumption: $E[Y_2(T,U) - Y_2(U,U)|\pi_1^*(H_1) = T, \pi_2^*(H_2(T)) = U]$ (average habit-formation effect for those assigned to (T,U)), $E[Y_2(T,T) - Y_2(U,T)|\pi_1^*(H_1) = T, \pi_2^*(H_2(T)) = T]$ (average learning effect for those assigned to (T,T)), and $E[Y_2(T,\pi_2^*(H_2(T))) - Y_2(T,\pi_2^*(H_2(U)))|\pi_1^*(H_1) = T]$ (average screening effect for treated in the first period). For example, regarding $E[Y_2(T,U) - Y_2(U,U)|\pi_1^*(H_1) = T, \pi_2^*(H_2(T)) = U]$, we cannot observe $Y_2(U,U)$ and $H_2(T)$ simultaneously; hence, the pair of $Y_2(U,U)$ and $H_2(T)$ is counterfactual.

Here, we propose a comprehensive approach to identifying these components. The proposed approach depends on the so-called *rank preservation/invariance assumption* (Chernozhukov and Hansen (2005)) on the state variables $S_2(d_1)$, where $S_2(d_1)$ comprises the two variables: The peak-time electricity consumption in the event period in the summer and the peak-time baseline electricity consumption in the winter.

Assumption 5.1. *(Rank invariance) Let K be the dimension of S_2, and we denote by $S_2^k(d_1)$ the k-th element of $S_2(d_1)$. For each k, there exists a function ψ_k and random variable ε_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 ψ_k is continuous and strictly increasing in ε_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 used by scholars such as Heckman et al. (1997), Chernozhukov and Hansen (2005), and Vuong and Xu (2017) to identify various causal parameters.[13] Note that $K = 2$ in our empirical context. Specifically, this assumption requires that the relative rank/quantile of the distribution of $S_{2,T,s_1}^k \equiv \psi_k(T, s_1, \varepsilon_k)$ is the same as that of the distribution of $S_{2,U,s_1}^k \equiv \psi_k(U, s_1, \varepsilon_k)$. If the error term is additive, as is often assumed (i.e., $S_2^k(d_1) = \psi_k^*(d_1, S_1) + \varepsilon_k$ for some real valued function ψ_k^*), then Assumption 5.1 is satisfied. Moreover, conditioning a sufficient set of pre-intervention information S_1 makes the rank-invariance plausible. In our empirical context, S_1 has a range of pre-treatment information,

[13] Chernozhukov and Hansen (2005) use the rank-invariance assumption to identify quantile treatment effects; Vuong and Xu (2017) use this assumption to identify individual treatment effects.

including baseline electricity consumption and household characteristics. Therefore, we do not consider this assumption to be restrictive in our empirical context.

For any random variables A and B, we denote by $Q_{A|B}(\cdot|b)$ the conditional quantile function of A given $B = b$; that is, $Q_{A|B}(\tau|b) = \inf\{a \in \mathbb{R} : P(A \leq a|b) \geq \tau\}$ for any $\tau \in [0,1]$. Let $F_{A|B}(\cdot|b)$ be the conditional distribution function of A given $B = b$. Under the rank-invariance condition (Assumption 5.1), the counterfactual state variable $S_2^k(d_1)$ is expressed as

$$S_2^k(d_1) = Q_{S_2^k(d_1)|S_1}(F_{S_2^k(d_1')|S_1}(S_2^k(d_1')|s_1)|S_1). \tag{10}$$

for any $d_1, d_1' \in \{U, T\}$. Therefore $S_2^k(T)$ and $S_2^k(U)$ have one-to-one mapping, which is called a *counterfactual mapping* in Vuong and Xu (2017). Under random assignment of (D_1, D_2), $F_{S_2^k(d_1')|S_1}$ and $Q_{S_2^k(d_1')|S_1}$ in equation (10) can be identified as

$$F_{S_2^k(d_1')|S_1}(\cdot|s_1) = F_{S_2^k|(D_1,S_1)}(\cdot|d_1, s_1), \tag{11}$$

$$Q_{S_2^k(d_1')|S_1}(\cdot|s_1) = Q_{S_2^k|(D_1,S_1)}(\cdot|d_1, s_1). \tag{12}$$

Therefore, through equations (10)– (12), we can identify (the distribution of) $S_2^k(d_1)$ even for those who do not take the intervention d_1 in the first period.

Building on these results, the following proposition shows the identifiability of the last three terms in equation (4) under the rank-invariance assumption and random assignment of the sequential intervention (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)* $(Y_2(d_1, d_2), S_2(d_1)) \perp\!\!\!\perp D_1|H_1$ *a.s. and (ii)* $Y_2(d_1, d_2) \perp\!\!\!\perp D_2|H_2$ *a.s. Let* $\widetilde{H}_2(d_1) = (S_1, d_1, \widetilde{S}_2(d_1))$ *with* $\widetilde{S}_2(d_1) = (\widetilde{S}_2^1(d_1), \ldots, \widetilde{S}_2^K(d_1))$ *and*

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

for $k = 1, 2, \ldots, K$. Then, given the optimal DTR π^*, the following hold:

$$
\begin{aligned}
&E[Y_2(T,U) - Y_2(U,U) | \pi_1^*(H_1) = T, \pi_2^*(H_2(T)) = U] \\
&= E\left[\frac{1\{D_1 = T\}}{P\left(D_1 = T | H_1\right)} \cdot \frac{1\{D_2 = U\}}{P\left(D_2 = U | H_2\right)} \cdot Y_2 \middle| \pi_1^*(H_1) = T, \pi_2^*(H_2) = U \right] \\
&- E\left[\frac{1\{D_1 = U\}}{P(D_1 = U | H_1)} \cdot \frac{1\{D_2 = U\}}{P(D_2 = U | H_2)} \cdot Y_2 \middle| \pi_1^*(H_1) = T, \pi_2^* \left(\widetilde{H}_2(T) \right) = U \right], \qquad (13) \\
&E[Y_2(T,T) - Y_2(U,T) | \pi_1^*(H_1) = T, \pi_2^*(H_2(T)) = T] \\
&= E\left[\frac{1\{D_1 = T\}}{P\left(D_1 = T | H_1\right)} \cdot \frac{1\{D_2 = T\}}{P\left(D_2 = T | H_2\right)} \cdot Y_2 \middle| \pi_1^*(H_1) = T, \pi_2^*(H_2) = T \right] \\
&- E\left[\frac{1\{D_1 = U\}}{P(D_1 = U | H_1)} \cdot \frac{1\{D_2 = T\}}{P(D_2 = T | H_2)} \cdot Y_2 \middle| \pi_1^*(H_1) = T, \pi_2^* \left(\widetilde{H}_2(T) \right) = T \right], \qquad (14)
\end{aligned}
$$

and

$$
\begin{aligned}
&E[Y_2\left(T, \pi_2^*\left(H_2(T)\right)\right) - Y_2\left(T, \pi_2^*\left(H_2(U)\right)\right) | \pi_1^*(H_1) = T] \\
&= E\left[\frac{1\{D_1 = T\}}{P(D_1 = T | H_1)} \cdot \frac{1\{D_2 = \pi_2^*(H_2)\}}{P(D_2 = \pi_2^*(H_2) | H_2)} \cdot Y_2 \middle| \pi_1^*(H_1) = T \right] \\
&- E\left[\frac{1\{D_1 = T\}}{P(D_1 = T | H_1)} \cdot \frac{1\left\{D_2 = \pi_2^*\left(\widetilde{H}_2(U)\right)\right\}}{P\left(D_2 = \pi_2^*\left(\widetilde{H}_2(U)\right) \middle| H_2\right)} \cdot Y_2 \middle| \pi_1^*(H_1) = T \right]. \qquad (15)
\end{aligned}
$$

Proof. See Appendix A.4. □

Conditions (i) and (ii) in Proposition 5.1 require (conditional) random assignment of D_1 and D_2, which are certainly satisfied in our experiment. Proposition 5.1 shows the identification of three components in the decomposition (4): the conditional averages of habit-formation effect, learning effect, and screening effect. The other components in equation (4) are straightforwardly identifiable and estimable under the random assignment assumption only.

The identification results in Proposition 5.1 constructively suggest a way to estimate each of the conditional averages of the habit-formation effect, learning effect, and screening effect. Let $\widehat{F}_{S_2^k|(D_1,S_1)}(\cdot|d_1,s_1)$ and $\widehat{Q}_{S_2^k|(D_1,S_1)}(\cdot|d_1,s_1)$ denote estimators of the conditional counterfactual distribution function $F_{S_2^k|(D_1,S_1)}(\cdot|d_1,s_1)$ and conditional counterfactual quantile function $Q_{S_2^k|(D_1,S_1)}(\cdot|d_1,s_1)$, respectively. To estimate $F_{S_2^k(D_1,S_1)}(\cdot|d_1,s_1)$, we apply the distribution regression with the standard logistic distribution function (e.g., Chernozhukov et al. (2013)) using the subsample with $D_1 = d_1$. To estimate $Q_{S_2^k(D_1,S_1)}(\cdot|d_1,s_1)$, we apply the linear quantile

regression using the subsample with $D_1 = d_1$. We estimate these with the test data.

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

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

Let $\widehat{S}_{i2}^{\text{test}}(d_1) = (\widehat{S}_{i2}^{1,\text{test}}, \widehat{S}_{i2}^{2,\text{test}})$ and $\widehat{H}_{i2}^{\text{test}}(d_1) = \left(S_{i1}^{\text{test}}, d_1, \widehat{S}_{i2}^{\text{test}}(d_1)\right)$. Then, using the test data, we estimate the average habit-formation effect for those assigned to (T,U) ($E[Y_2(T,U) - Y_2(U,U)|\pi_1^*(H_1) = T, \pi_2^*(H_2(T)) = U]$), the average learning effect for those assigned to (T,T) ($E[Y_2(T,T) - Y_2(U,T)|\pi_1^*(H_1) = T, \pi_2^*(H_2(T)) = T]$), and the average screening effect for those treated in the first period ($E[Y_2\left(T, \pi_2^*\left(H_2(T)\right)\right) - Y_2\left(T, \pi_2^*\left(H_2(U)\right)\right)|\pi_1^*(H_1) = T]$), respectively, as follows:

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

5.2 Estimation Results

Table 6 shows the estimation results of each component comprising the decomposition (4). All five components positively contribute to the welfare gain of the optimal DTR π^*. The total contribution of all components is estimated as 1613.1 JPY, being close to the estimate of welfare gain of π^* presented in Table 2 (1684.3 JPY). The first-period intervention has a welfare contribution of 214.3 JPY, implying that 13% of welfare gain comes from the contribution to the first-period outcome. The two behavioral concepts, habit formation and learning , account for 287.4 JPY (18%) and 186.4(12%) of the welfare gain, respectively. The results also show that the screening effect also positively contributes to the welfare gain by 361.5 JPY

(22%), implying that drawing the information $H_2(T)$ is useful to improve the welfare for those assigned to $d_1 = T$ by the first-period optimal policy π_1^*.

[Table 6 about here]

Appendix A.5 further investigates the mechanism of the welfare improvement by the optimal DTP.

6 Conclusion

This study introduced a framework for designing optimal dynamic targeting strategies that maximize social welfare gains from dynamic policy interventions, using experimental or quasi-experimental data. We theoretically demonstrate that dynamic targeting can surpass static targeting via several key mechanisms, namely learning, habit formation, and screening effects.

We apply this methodology to an RCT of a residential energy rebate program. Our empirical findings reveal that dynamic targeting significantly outperforms conventional static targeting, thereby enhancing the social welfare benefits derived from the energy rebate program. We identify considerable heterogeneity in the learning, habit formation, and screening effects across households. This paper illustrates how our approach leverages this heterogeneity to devise optimal dynamic targeting strategies.

References

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*.

ASSUNÇÃO, J., R. MCMILLAN, J. MURPHY, AND E. SOUZA-RODRIGUES (2023): "Optimal environmental targeting in the amazon rainforest," *The Review of Economic Studies*, 90, 1608–1641.

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

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.

CAGALA, T., U. GLOGOWSKY, J. RINCKE, AND A. STRITTMATTER (2021): "Optimal Targeting in Fundraising: A Causal Machine-Learning Approach," *arXiv preprint arXiv:2103.10251*.

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.

CHRISTENSEN, P., P. FRANCISCO, E. MYERS, H. SHAO, AND M. SOUZA (2021): "Energy Efficiency Can Deliver for Climate Policy: Evidence from Machine Learning-Based Targeting," *Working Paper*.

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.

GERARDEN, T. D. AND M. YANG (2023): "Using targeting to optimize program design: evidence from an energy conservation experiment," *Journal of the Association of Environmental and Resource Economists*, 10, 687–716.

HAN, S. (2023): "Optimal dynamic treatment regimes and partial welfare ordering," *Journal of the American Statistical Association*, 1–11.

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, 2 ed.

HECKMAN, J. J., J. E. HUMPHRIES, AND G. VERAMENDI (2016): "Dynamic treatment effects," *Journal of econometrics*, 191, 276–292.

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.

HECKMAN, J. J. AND S. NAVARRO (2007): "Dynamic discrete choice and dynamic treatment effects," *Journal of Econometrics*, 136, 341–396.

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 (2023): "Choosing Who Chooses: Selection-driven targeting in energy rebate programs," *NBER working paper*.

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.*

IDA, T. AND Y. USHIFUSA (2017): "Cost-Benefit Analysis of Price-based Residential Demand Response," *Proceedings of the Japan Joint Automatic Control Conference*, 60, 304–307.

ITO, K. (2015): "Asymmetric Incentives in Subsidies: Evidence from a Large-Scale Electricity Rebate Program," *American Economic Journal: Economic Policy*, 7, 209–37.

ITO, K., T. IDA, AND M. TAKANA (2023): "Selection on Welfare Gains: Experimental Evidence from Electricity Plan Choice," *American Economic Review*, 113, 2937–73.

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.

JOHNSON, M. S., D. I. LEVINE, AND M. W. TOFFEL (forthcoming): "Improving regulatory effectiveness through better targeting: Evidence from OSHA," *American Economic Journal: Applied Economics*.

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," .

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. (2022): "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.

MURAKAMI, K., H. SHIMADA, Y. USHIFUSA, AND T. IDA (2022): "Heterogeneous treatment effects of nudge and rebate: Causal machine learning in a field experiment on electricity conservation," *International Economic Review*, 63, 1779–1803.

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

——— (2005): "A Generalization Error for Q-learning," *Journal of Machine Learning Research*, 6, 1073–1097.

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.

REISS, P. C. AND M. W. WHITE (2008): "What changes energy consumption? Prices and public pressures." *RAND Journal of Economics*, 39, 636–663.

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. (2021): "Estimation of Optimal Dynamic Treatment Assignment Rules under Policy Constraints," *arXiv preprint arXiv:2106.05031*.

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.

WATKINS, C. J. AND P. DAYAN (1992): "Q-learning," *Machine learning*, 8, 279–292.

WOLAK, F. A. (2006): "Residential Customer Response to Real-Time Pricing: the Anaheim Critical-Peak Pricing Experiment," *Working Paper*.

——— (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.

Tables

Table 1: Summary Statistics and Balance Check

	Sample mean by group [standard deviation]				Difference in sample means
	(U,U)	(U,T)	(T,U)	(T,T)	p-value
Peak hour usage (2020 summer, Wh)	201 [145]	200 [136]	196 [136]	198 [136]	0.910
Pre-peak hour usage (2020 summer, Wh)	189 [143]	184 [130]	183 [137]	182 [130]	0.853
Post-peak hour usage (2020 summer, Wh)	311 [175]	311 [171]	308 [164]	305 [163]	0.893
Peak hour usage (2020 winter, Wh)	311 [194]	309 [170]	304 [179]	306 [170]	0.913
Pre-peak hour usage (2020 winter, Wh)	171 [117]	171 [102]	169 [112]	166 [102]	0.854
Post-peak hour usage (2020 winter, Wh)	287 [198]	295 [198]	280 [203]	287 [192]	0.664
Number of people at home (1 PM - 5 PM)	1.31 [1.04]	1.32 [0.96]	1.31 [1.04]	1.34 [1.01]	0.956
Number of people at home (5 PM - 9 PM)	2.57 [1.29]	2.48 [1.20]	2.47 [1.23]	2.51 [1.20]	0.483
Self-efficacy in energy conservation (1-5 scale)	3.44 [0.84]	3.44 [0.86]	3.47 [0.86]	3.44 [0.82]	0.935
Household income (JPY 10,000)	651 [400]	639 [387]	614 [393]	606 [333]	0.120

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. Column 5 shows the p-values of the F-test for the difference in sample averages between the four groups. 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 ¢=1 JPY in the summer of 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 (0.0)	100.0%	0.0%	0.0%	0.0%
100% (T,U)	311.8 (378.4)	0.0%	100.0%	0.0%	0.0%
100% (U,T)	470.8 (457.5)	0.0%	0.0%	100.0%	0.0%
100% (T,T)	463.9 (452.2)	0.0%	0.0%	0.0%	100.0%
Static targeting I ($\pi^*_{S(I)}$)	770.6 (283.7)	45.6%	0.0%	0.0%	54.4%
Static targeting II ($\pi^*_{S(II)}$)	845.3 (348.9)	3.1%	31.3%	41.5%	24.0%
Dynamic targeting (π^*)	1684.3 (303.1)	19.5%	22.9%	25.6%	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 (I) ($\pi^*_{S(I)}$), static targeting (II) ($\pi^*_{S(II)}$), and dynamic targeting (π^*). 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 ¢ = 1 JPY, the exchange rate in the summer of 2020.

Table 3: Comparisons of Alternative Policies

	Difference in welfare gains	p-value
Dynamic targeting (π^*) vs. 100% (T,U)	1365.7 (309.1)	0.000
Dynamic targeting (π^*) vs. 100% (U,T)	1546.5 (328.7)	0.000
Dynamic targeting (π^*) vs. 100% (T,T)	1397.7 (319.8)	0.000
Dynamic targeting (π^*) vs. Static targeting I ($\pi^*_{S(I)}$)	913.8 (269.2)	0.000
Dynamic targeting (π^*) vs. Static targeting II ($\pi^*_{S(II)}$)	839.1 (287.8)	0.002

Notes: This table compares welfare gains from each policy. For each row, the column "Difference in Welfare Gains" shows the estimated welfare gain of the policy on the left-hand side (W_L) relative to the policy on the right-hand side (W_R) in JPY per household, with its standard error in parenthesis. The column "p-value" gives the p-value for the null hypothesis: $H_0 : W_L \leq W_R$. The monetary unit is given as 1 ¢ = 1 JPY, the exchange rate in the summer of 2020.

Table 4: Welfare Differences in Stages 1 and 2

	Welfare difference	
	Stage 1	Stage 2
Dynamic targeting (π^*) vs. 100% (U,U)	298.4	1395.0
	(103.1)	(255.6)
Dynamic targeting (π^*) vs. 100% (T,U)	64.1	1301.6
	(89.8)	(268.6)
Dynamic targeting (π^*) vs. 100% (U,T)	289.4	1257.1
	(103.1)	(285.7)
Dynamic targeting (π^*) vs. 100% (T,T)	64.1	1333.5
	(89.8)	(284.7)
Dynamic targeting (π^*) vs. Static targeting I ($\pi^*_{S(I)}$)	105.7	808.0
	(81.4)	(237.9)
Dynamic targeting (π^*) vs. Static targeting II ($\pi^*_{S(II)}$)	245.3	593.8
	(90.6)	(251.3)

Notes: This table compares welfare gains from each policy. For each row, the columns "Stage 1" and "Stage 2" show the estimated welfare gains from the stages 1 and 2 of the policy on the left-hand side (W_L) relative to the policy on the right-hand side (W_R) in JPY per household per season, with its standard error in parenthesis. The monetary unit is given as 1 ¢ = 1 JPY, the exchange rate in the summer of 2020.

Table 5: Welfre Comparison in Eeach Group Assinged by the Dynamic Targeting

Assigned group	Comparison intervention			
	(U,U)	(T,U)	(U,T)	(T,T)
(U,U)	0.0	-271.7	601.3	1097.4
	(0.0)	(1012.1)	(936.5)	(1068.9)
(T,U)	842.7	0.0	1773.1	994.5
	(904.3)	(0.0)	(979.3)	(901.2)
(U,T)	1258.2	2051.7	0.0	1181.9
	(484.0)	(638.1)	(0.0)	(565.0)
(T,T)	1819.7	1062.2	995.1	0.0
	(528.9)	(444.0)	(492.9)	(0.0)

Notes: This table shows the each estimate of counterfactual welfare difference $WD^{\hat{\pi}^*}_{(d_1,d_2),(d'_1,d'_2)}$ for each group $(d_1,d_2) \in \{U,T\}^2$ assigned by the estimated DTP $\hat{\pi}^*$ (row) relative to each counterfactual intervention ($(d'_1,d'_2) \in \{U,T\}^2$) (colmun), with its standard error is in parenthesis. Specifically, for "assigned group"' being (d_1,d_2) and "comparison intervention"' being (d'_1,d'_2), the corresponding celles show estimate of $WD^{\hat{\pi}^*}_{(d_1,d_2),(d'_1,d'_2)}$ along with its stanard error. Each welfare difference is estimated with the test data. The monetary unit is given as 1 ¢ = 1 JPY, the exchange rate in the summer of 2020.

Table 6: Decomposition of Welfare Gain for the Optimal Dynamic Targeting π^*

	Conditional effect	Fraction	Welfre contribution
1st-stage treatment effect	390.2	0.55	214.3
	(187.5)	(0.00)	(103.0)
2nd-stage treatment effect	996.7	0.57	563.5
	(350.6)	(0.00)	(198.3)
Habit formation effect	1282.0	0.22	287.4
	(897.0)	(0.01)	(184.4)
Learning effect	573.6	0.32	186.4
	(419.1)	(0.01)	(128.8)
Screening effect	658.3	0.55	361.5
	(178.4)	(0.01)	(98.0)
Total effect			1613.1
			(397.8)

Notes: This table shows estimation results for the decomposition (4). The column "Conditional effect" shows the estimates of conditional average effects that appear in equation (4), 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 (4). 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 (4). The standard errors are in parentheses. The monetary unit is given as 1 ¢ = 1 JPY, the exchange rate in the summer of 2020.

A Online Appendix

A.1 Proof of Proposition 2.1

We prove a general version of the proposition by letting the benchmark policy for the welfare comparison be a uniform assignment of $(d_1, d_2) \in \{U, T\}^2$ and DTP π be arbitrary. In what follows, we let d_1' and d_2' be the other treatment status of d_1 and d_2, respectively (e.g., if $d_1 = U$ and $d_2 = U$, $d_1' = T$ and $d_2' = T$).

Consider first the following decomposition that follows from the law of total probabilities regarding the events of $\{\pi_1(H_1) = d_1\}$ and its complement:

$$
\begin{aligned}
& W(\pi) - W(d_1, d_2) \\
=& E[Y_2(d_1, \pi_2(H_2(d_1))) - Y_2(d_1, d_2) | \pi_1(H_1) = d_1] \Pr(\pi_1(H_1) = d_1) && (16) \\
& + E[Y_1(d_1') - Y_1(d_1) | \pi_1(H_1) = d_1'] \Pr(\pi_1(H_1) = d_1') && (17) \\
& + E[Y_2(d_1', \pi_2(H_2(d_1'))) - Y_2(d_1, d_2) | \pi_1(H_1) = d_1'] \Pr(\pi_1(H_1) = d_1'). && (18)
\end{aligned}
$$

Given DTR π and (d_1, d_2), we define the following subsets in the population:

$$
\begin{aligned}
A_1 &= \{\pi_1(H_1) = d_1\}, \\
A_1' &= \{\pi_1(H_1) = d_1'\}, \\
A_2 &= \{\pi_2(H_2(d_1)) = d_2\}, \\
A_2' &= \{\pi_2(H_2(d_1)) = d_2'\}.
\end{aligned}
$$

Note that Term (16) is nonzero if and only if $\pi_2(H_2(d_1))) = d_2'$. Hence, it can be written as

$$
(16) = E[W_2(d_1, d_2') - W_2(d_1, d_2) | A_2' \cap A_1] \cdot \Pr(A_2' \cap A_1). \tag{19}
$$

Term (17) can be seen as the average treatment effect for the treated in period 1 multiplied by the proportion of the treated.

Next, consider decomposing Term (18) as

$$(18) = E[Y_2(d_1', \pi_2(H_2(d_1'))) - Y_2(d_1', \pi_2(H_2(d_1)))|A_1'] \Pr(A_1') \tag{20}$$

$$+ E[Y_2(d_1', \pi_2(H_2(d_1))) - Y_2(d_1, \pi_2(H_2(d_1)))|A_1'] \Pr(A_1') \tag{21}$$

$$+ E[Y_2(d_1, \pi_2(H_2(d_1))) - Y_2(d_1, d_2)|A_1'] \Pr(A_1'). \tag{22}$$

Similar (19), term (22) is nonzero if and only if $\pi_2(H_2(d_1))) = d_2'$. Therefore, we have

$$(16) = E[Y_2(d_1, d_2') - Y_2(d_1, d_2)|A_2' \cap A_1'] \cdot \Pr(A_2' \cap A_1'). \tag{23}$$

The sum of terms (19) and (23) yields

$$(19) + (22) = E[Y_2(d_1, d_2') - Y_2(d_1, d_2)|A_2'] \cdot \Pr(A_2'). \tag{24}$$

We can represent term (21) as

$$(21) = E[Y_2(d_1', d_2) - Y_2(d_1, d_2)|A_2 \cap A_1'] \Pr(A_2 \cap A_1') + E[Y_2(d_1', d_2') - Y_2(d_1, d_2')|A_2' \cap A_1'] \Pr(A_2' \cap A_1'). \tag{25}$$

Note that term (20) corresponds to the conditional average screening effect of d_1' relative to d_1 at period 2 assignment policy π_2, fixing period 1 treatment to d_1'.

Combining Terms (17), (24), (25), and (20) and setting $d_1 = U$, $d_2 = T$, and $\pi = \pi^*$, we obtain the current proposition. □

A.2 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:

$$Q_{ih} = \beta_T T_{ih} + \lambda_i + \theta_h + \epsilon_{ih}, \tag{26}$$

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.[14] 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 λ_i and time fixed effects θ_h in each 30-minute interval to control for time-specific shocks such as weather. Given that $D_t = \{U, T\}$ is randomly assigned, β_T provides the ATE of $D_t = T$ relative to $D_t = U$. We use cluster standard errors at the household level.

Table A.1 presents the estimation results for equation 26. 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.1 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:

$$Q_{ih} = \beta_{UT} UT_{ih} + \beta_{TU} TU_{ih} + \beta_{TT} TT_{ih} + \lambda_i + \theta_h + \epsilon_{ih}, \tag{27}$$

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 β_{TU} represents the extent to which the effect of the treatment received in the first period is sustained in the second period. Thus,

[14] 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 the household-specific time-invariant variation.

this effect can be regarded as a habit-formation effect. We also regard a negative difference between the coefficients β_{TT} and β_{UT} as a learning effect and a positive difference as a habituation effect.

[Table A.2 about here]

Table A.2 presents the estimation results for equation 27. 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 DTP presented in Section 2 may increase welfare gains from the policy. The remaining columns of Table A.2 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 PM - 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.3 Artificial Test Data

We describe our method for generating artificial test data in detail. First, to motivate our construction of artificial test data, we briefly discuss why bias in point estimates of welfare gain is introduced when one uses the same data to learn an optimal policy and estimate its welfare gain. Simply, the bias occurs given the noise in the observed electricity consumption Q_1 and Q_2, which are random components in 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:

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

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 necessary for learning an optimal policy, a learning algorithm inevitably responds to the noise term and overfits the training sample at hand. Therefore, when one evaluates the welfare performance of the estimated policy on the same training sample, the welfare estimate is biased upward because the policy also fits the noise term. Hence, if we replace the noise term in the training sample with another independent noise sample, we can eliminate the bias from the estimate of welfare performance.

Motivated by this observation, we generate test data $\mathcal{S}^{\text{test}} = \{S_{i1}, S_{i2}, Y_{i1}^{\text{test}}, Y_{i2}^{\text{test}}, D_{i1}, D_{i2} : i = 1, \ldots, n\}$ by generating another noise sample. Here, $Y_{it}^{\text{test}} := b \cdot Q_{it}^{\text{test}} - a \cdot 1\{D_{it} = T\}$, with Q_{it}^{test} being electricity consumption in the test data. For Q_{i1}^{test} and Q_{i2}^{test}, we use the following procedure to generate artificial data:

1. For $d_1 \in \{U,T\}$ and subsample $\{i = 1, \ldots, n : D_{i1} = d_1\}$,

   (a) Estimate the conditional expectation function of potential electricity usage in the first period $E[Q_1(d_1)|H_1 = h_1]$ and calculate residuals $\hat{\epsilon}_{i1,d_1} = Q_{i1} - \hat{E}[Q_1(d_1)|H_1 = H_{i1}]$, where

$\hat{E}[Q_1(d_1)|H_1 = h_1]$ is the regression fitted value and $\hat{\epsilon}_{i1,d_1}$ is an estimate of the noise term ϵ_{i1,d_1}.

(b) Estimate conditional variance of the regression residuals $E[\epsilon^2_{i1,d_1}|H_1 = H_{i1}]$ by regressing $\hat{\epsilon}^2_{1i,d_1}$ on H_{i1} and calculate $\hat{\sigma}^2_{i1,d_1} = \hat{E}[\epsilon^2_{1,d_1}|H_1 = H_{i1}]$.

2. For $(d_1, d_2) \in \{U, T\}^2$ and subsample $\{i : (D_{i1}, D_{i2}) = (d_1, d_2)\}$,

(a) Estimate the conditional expectation function of potential electricity usage in the second period $E[Q_2(d_1, d_2)|H_2 = h_2]$ and calculate residuals $\hat{\epsilon}_{i2,(d_1,d_2)} = Q_{i2} - \hat{E}[Q_2(d_1, d_2)|H_2 = H_{i2}]$, where $\hat{E}[Q_2(d_1, d_2)|H_2 = h_2]$ is the regression fitted value, and $\hat{\epsilon}_{i2,(d_1,d_2)}$ is an estimate of the noise term $\epsilon_{i2,(d_1,d_2)}$.

(b) Estimate conditional variance of the regression residuals $E[\epsilon^2_{i2,(d_1,d_2)}|H_2 = H_{i2}]$ by regressing $\hat{\epsilon}^2_{i2,(d_1,d_2)}$ on H_{i2} and calculate $\hat{\sigma}^2_{i2,(d_1,d_2)} = \hat{E}[\epsilon^2_{2,(d_1,d_2)}|H_2 = H_{i2}]$.

(c) Estimate 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_{i1} and calculate $\hat{\sigma}_{i12,(d_1,d_2)} = \hat{E}[\epsilon_{1,d_1}\epsilon_{2,(d_1,d_2)}|H_1 = H_{i1}]$.

(d) For each $i \in \{i : (D_{i1}, D_{i2}) = (d_1, d_2)\}$, estimate the covariance matrix as

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

Sample $\{(\tilde{\epsilon}_{i1}, \tilde{\epsilon}_{i2}) : i \in \{i : (D_{i1}, D_{i2}) = (d_1, d_2)\}\}$ iid from the empirical distribution of the standardized residuals $\{(\hat{\epsilon}_{i1,d_1}, \hat{\epsilon}_{i2,(d_1,d_2)})\widehat{\Sigma}_i^{-1/2} : i \in \{i : (D_{i1}, D_{i2}) = (d_1, d_2)\}\}$ and calculate $\epsilon_i^{\text{test}} = (\epsilon_{i1}^{\text{test}}, \epsilon_{i2}^{\text{test}})^T = \widehat{\Sigma}_i^{-1/2}(\tilde{\epsilon}_{i1}, \tilde{\epsilon}_{i2})^T$.

(e) Construct $(Q_{i1}^{\text{test}}, Q_{i1}^{\text{test}})^T = (\hat{E}[Q_1(d_1)|H_1 = H_{i1}], \hat{E}[Q_2(d_1, d_2)|H_2 = H_{i2}])^T + \epsilon_i^{\text{test}}$.

In this procedure, we estimate the conditional expectation functions of $Q_1(d_1)$ and $Q_2(d_1, d_2)$ using random forests (Friedberg, Tibshirani, Athey, and Wager, 2021; Wager and Athey, 2018).

A.4 Proof of Proposition 5.1

By the same argument in the proof of Lemma 1 in Vuong and Xu (2017), under Assumption 5.1, equation (10) holds a.s. for any $d_1, d_1' \in \{U, T\}$. Under the conditions (i) and (ii) of the random assignment of

(D_1, D_2), $Q_{S_2^k|(D_1,S_1)}(\cdot|d_1, S_1) = Q_{S_2^k(d_1)|S_1}(\cdot|S_1)$ and $F_{S_2^k|(D_1,S_1)}(\cdot|d_1, S_1) = F_{S_2^k(d_1)|S_1}(\cdot|S_1)$ hold a.s. Hence, we have $\widetilde{S}_2^k(d_1) = S_2^k(d_1)$ a.s.

Regarding the conditional average of the habit-formation effect in equation (4), letting $\mathcal{A} := \{h_2 : \pi_1^*(h_1) = T, \pi_2^*(h_2) = U\}$,

$$
\begin{aligned}
&E\left[\frac{1\{D_1 = T\}}{P\left(D_1 = T|H_1\right)} \cdot \frac{1\{D_2 = U\}}{P\left(D_2 = U|H_2\right)} \cdot Y_2 \middle| \pi_1^*(H_1) = T, \pi_2^*(H_2) = U\right] \\
&= E\left[\frac{1\{D_1 = T\}}{P\left(D_1 = T|H_1\right)} \cdot \frac{1\{D_2 = U\}}{P\left(D_2 = U|H_2\right)} \cdot Y_2(T,U) \middle| H_2 \in \mathcal{A}\right] \\
&= E\left[\frac{1\{D_1 = T\}}{P\left(D_1 = T|H_1\right)} \cdot E\left[\frac{1\{D_2 = U\}}{P\left(D_2 = U|H_2\right)} \middle| H_2, H_2 \in \mathcal{A}\right] \cdot E\left[Y_2(U,T)|H_2, H_2 \in \mathcal{A}\right] \middle| H_2 \in \mathcal{A}\right] \\
&= E\left[\frac{1\{D_1 = T\}}{P\left(D_1 = T|H_1\right)} \cdot Y_2(U,T) \middle| H_2 \in \mathcal{A}\right] \\
&= E\left[Y_2(U,T)|H_2 \in \mathcal{A}\right], \qquad (28)
\end{aligned}
$$

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

$$
\begin{aligned}
&E\left[\frac{1\{D_1 = U\}}{P\left(D_1 = U|H_1\right)} \cdot \frac{1\{D_2 = T\}}{P\left(D_2 = T|H_2\right)} \cdot Y_2 \middle| \pi_1^*(H_1) = T, \pi_2^*\left(\widetilde{H}_2(T)\right) = U\right] \\
&= E\left[\frac{1\{D_1 = U\}}{P\left(D_1 = U|H_1\right)} \cdot \frac{1\{D_2 = T\}}{P\left(D_2 = T|H_2\right)} \cdot Y_2 \middle| \pi_1^*(H_1) = T, \pi_2^*\left(H_2(T)\right) = U\right] \\
&= E\left[Y_2(U,T)|\pi_1^*(H_1) = T, \pi_2^*\left(H_2(T)\right) = U\right], \qquad (29)
\end{aligned}
$$

where the second line follows from the result $\widetilde{S}_2^k(d_1) = S_2^k(d_1)$ a.s., and the last line follows from the same argument to derive (28). Combining these results yields (13). The results (14) and (13) can be also shown by a similar argument. □

A.5 Learning and Habit Formation Effects

Sections 2.3.1 and 2.3.2 discuss that learning and habit formation are behavior states relevant to the effectiveness of the sequential intervention of the rebate program. To improve social welfare, a targeting policy should utilize learning and habit-formation effects for the sequential treatment choice. For example, for consumers with a learning effect (i.e., $Y_2(T,T) > Y_2(U,T)$), the intervention (U,T) should not be provided given its lower welfare contribution in period 2. We investigate how optimal dynamic targeting

exploits these effects.

[Table A.3 about here]

Table A.3 presents the estimation results of the learning, habit formation, second-period-treatment, and full-intervention effects for each group $(d_1, d_2) \in \{U, T\}^2$ assigned by the optimal DTP π^*. In this table, we estimate the welfare gain for the second-period outcome to quantify each effect in Columns 1-4. Column 1 shows the average learning effect, $E[Y_2(T,T) - Y_2(U,T)]$, for each of the assigned groups. For instance, we find an economically and statistically significant positive learning effect (995.1 JPY) for those assigned to (T,T). This finding implies that the positive learning effect is among the reasons we want to assign this group to (T,T). However, we find a negative learning effect (i.e., a fatigue or decaying effect) for those assigned to (U,T). This result suggests that the fatigue effect is among the reasons we want to assign this group to (U,T). Column 2 shows the habit formation effect, $E[Y_2(T,U) - Y_2(U,U)]$, for each of the assigned groups. We find an economically significant positive habit-formation effect for those assigned to (T,U) and (T,T), although these estimates are noisy and statistically insignificant from zero.

A.6 Additional Tables

Table A.1: Average Treatment Effects

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

Notes: This table shows the estimation results for equation 26. 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.2: 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 27 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 the summer of 2020.

Table A.3: Learning, Habit Formation, Second-Stege-Treatment, and Full-Intervention Effects

Assigned group	Learning $E[Y_2(T,T) - Y_2(U,T)]$	Habit formation $E[Y_2(T,U) - Y_2(U,U)]$	2nd stage treatment $E[Y_2(U,T) - Y_2(U,U)]$	Full intervention $E[Y_2(T,T) - Y_2(U,U)]$
(U,U)	-496.1	271.7	-601.3	-1097.4
	(1040.7)	(1012.1)	(936.5)	(1068.9)
(T,U)	778.7	842.7	-930.4	-151.7
	(944.7)	(904.3)	(799.4)	(866.3)
(U,T)	-1181.9	-793.5	1258.2	76.3
	(565.0)	(688.6)	(484.0)	(620.7)
(T,T)	995.1	757.6	824.6	1819.7
	(492.9)	(591.7)	(497.3)	(528.9)

Notes: This table shows the estimates of averages of the four effects (learning, habit formation, second-stage-treatment, and full intervention effects) for each of the four intervention groups $(d_1, d_2) \in \{U,T\}^2$ assigned by the estimated DTP $\hat{\pi}^*$ and for the whole population with these standard errors in prarentheses. For example, the first row presents the estimates of averages of the four effects for the subpopulation that is assigned to (U,U) by the estimated DTP $\hat{\pi}$, where each average effect is estimated with the test data. The last row shows the estimates of averages of the four effects for the whole population. The monetary unit is given as 1 ¢ = 1 JPY, the exchange rate in the summer of 2020.