Policy-Aware Experimentation: Strategic Sampling for Optimized Targeting Policies

Yi-Wen Chen,[1*] Eva Ascarza,[2] Oded Netzer[3]

[1] Columbia Business School, Columbia University
[2] Harvard Business School, Harvard University
[3] Columbia Business School, Columbia University

June 21, 2025

Abstract

Firms often rely on randomized experiments to estimate customer-level treatment effects for targeting policies. Standard "test-then-learn" approaches typically sample customers uniformly to optimize estimation accuracy but ignore economic objectives, leading to statistically sound yet economically suboptimal targeting policies. We propose a novel experimental design — *policy-aware experimentation* — that directly incorporates firm's profit-maximizing objective into the sampling strategy. Specifically, we introduce *expected profit loss (EPL)*, a criterion that prioritizes sampling customers whose treatment effect estimation errors most likely affect profitability. We prove that allocating experimental resources based on EPL achieves near-optimality with theoretical guarantees, and develop a sequential sampling algorithm that prioritizes customers with highest EPLs for practical implementation.

Using simulations and two empirical applications, we show that our approach yields more profitable targeting policies than existing methods. Across both applications, our approach improves targeting performance by 5% to 10% in profit terms relative to standard methods, and achieves comparable outcomes with up to 80% fewer experimental samples, highlighting substantial gains in both targeting effectiveness and data efficiency.

Managerially, our results underscore the importance of aligning experimental design with business objectives. By focusing data collection on customers who matter most for decision quality, firms can develop more effective targeting policies without increasing experimental costs.

Key words: Policy learning, marketing interventions, targeted policies, experimentation, active learning, heterogeneous treatment effect

*Yi-Wen Chen (YChen26@gsb.columbia.edu) is the corresponding author. She is a PhD Candidate at Columbia Business School. Eva Ascarza is the Jakurski Family Associate Professor of Business Administration at Harvard Business School. Oded Netzer is the Arthur J. Samberg Professor of Business at Columbia Business School. The authors thank participants at the CBS "Quant Lab", the Customer Intelligence Lab at Harvard, 2025 ISMS Marketing Science conference, CODE@MIT conference, and the 42^{nd} University of Houston Doctoral Symposium for valuable feedbacks.

1. Introduction

Firms today have unprecedented access to detailed consumer information, enabling the design of highly customized targeting strategies. To develop these strategies, firms increasingly rely on randomized experiments to estimate how customers respond to marketing interventions and use these estimates to build data-driven targeting policies (e.g., Ascarza 2018; Simester et al. 2020; Yang et al. 2023). Targeting decisions are particularly critical when interventions involve non-trivial costs or the risk of cannibalizing existing revenues, as is often the case with promotions or personalized incentives.

The standard approach to targeting-based experimentation in marketing — often referred to as "test-then-learn" — typically follows three steps: (1) firms randomly sample customers to form experimental groups, (2) estimate the conditional average treatment effects (CATEs) (or, in some cases, estimate directly a policy based on those predicted CATEs), and (3) apply a decision rule to target remaining customers based on the predicted CATEs (e.g., Lemmens and Gupta 2020; Ellickson et al. 2022; Huang and Ascarza 2024). While this approach is statistically sound, the sampling strategy in the first step is typically designed to ensure representativeness and to reduce estimation error uniformly across the population — characteristics that are well-desired when the objective is to estimate the treatment effects. However, the firm's ultimate objective is not to precisely estimate CATEs for its customers, but rather to maximize profitability (or any other outcome of interest) through effective targeting. In other words, this sampling strategy may be misaligned with the firm's objective.

This misalignment can lead firms to allocate experimental resources to customers whose outcomes are largely irrelevant for determining optimal targeting policies, while undersampling those whose data could significantly impact profitability. In turn, firms may deploy targeting policies that are statistically sound but economically suboptimal. Recent research has begun to distinguish between predictive accuracy and decision quality in marketing experimentation (Fernández-Loría and Provost 2022), but questions remain about how to design experiments that prioritize the right customers for learning optimal targeting policies.

In this paper, we build on ideas from adaptive experimentation, active learning, and decision-aware learning to propose a novel approach that explicitly aligns the firm's profit-maximizing objective with its sampling strategy — we term this *policy-aware experimentation*. Specifically, we introduce a sequential experimental design guided by a profit-based sampling criterion — *expected* *profit loss (EPL)*—which prioritizes customers whose estimation errors are most likely to distort targeting decisions and reduce profitability. We theoretically show that the EPL approach achieves near-optimality with a provable performance guarantee, and develop an estimation strategy leveraging the power of Bayesian inference for uncertainty quantification based on the Causal Forest framework (Wager and Athey 2018; Athey et al. 2019b). By sampling the most consequential customers more intensively, our method minimizes prediction errors in treatment effect estimation where they matter most for targeting profitability.

We evaluate the benefits of policy-aware experimentation in two real-world applications—a telecommunications reactivation campaign and a Starbucks promotional offer—as well as extensive simulation studies. We empirically demonstrate that our method yields more profitable targeting policies than conventional sampling designs, including standard test-then-learn, uncertainty sampling, and a leading adaptive design (e.g., Kato et al. 2024). Across these empirical settings, our approach improves profitability by 5% to 10% relative to standard methods, and achieves comparable outcomes with up to 80% fewer experimental samples. These results highlight both the economic relevance and the sample efficiency of our method, especially in settings where experimental budgets are limited.

These gains are most pronounced in settings where only a small fraction of customers are positioned near the decision threshold. Our method's advantage over alternative sampling strategies arises in settings where there is misalignment between intervention costs and the distribution of customer responsiveness, specifically when: (1) the intervention is broadly harmful (e.g., Ascarza et al. 2016); (2) the intervention is costly with uncertain returns (e.g., Lemmens and Gupta 2020; Simester et al. 2022); or (3) the intervention risks cannibalizing revenues that would have occurred naturally (e.g. Anderson and Simester 2004; Ascarza 2018; Yang et al. 2023). Finally, we show that our sequential sampling method maintains its effectiveness in a simplified two-stage design, reducing both practical and technical barriers to implementation.

This research makes three key contributions. First, we introduce a new sampling criterion, expected profit loss (EPL), that theoretically aligns with the firm's profit-maximizing objective and achieves near-optimality. Second, we demonstrate that a policy-aware experimental design can substantially improve targeting outcomes even without increasing sample size. Third, we provide a practical and managerially relevant solution that allows firms to conduct more efficient experiments by strategically focusing on the most consequential customers.

2. Problem Formulation

Firms aiming to optimize marketing targeting decisions often rely on predictions of how individual customers will respond to an intervention. However, not all customers contribute equally to learning an effective targeting policy. Some are more consequential than others, particularly those for whom small errors in prediction can translate into large differences in profitability.

Consider a firm that is deciding whether to send a $2 coupon to each customer. The optimal decision rule would be to target only customers whose incremental spending from receiving the coupon exceeds $2. In practice, however, the true incremental spending — often referred to as the conditional average treatment effect (CATE) — is not observed and must be estimated from data. As a result, targeting decisions are based on predictions, which inevitably include prediction error.

To illustrate the implications of this error, consider three customers, A, B, and C, in Figure 1, who should be targeted (as their true CATE is above the decision threshold). The three customers also have the same prediction error. For customer C, whose actual incremental spending is significantly above $2 (or, more broadly, any customer whose true CATE is far from the decision threshold), moderate errors in CATE estimation are relatively inconsequential, as the firm can still make the correct targeting decision despite some degree of error.

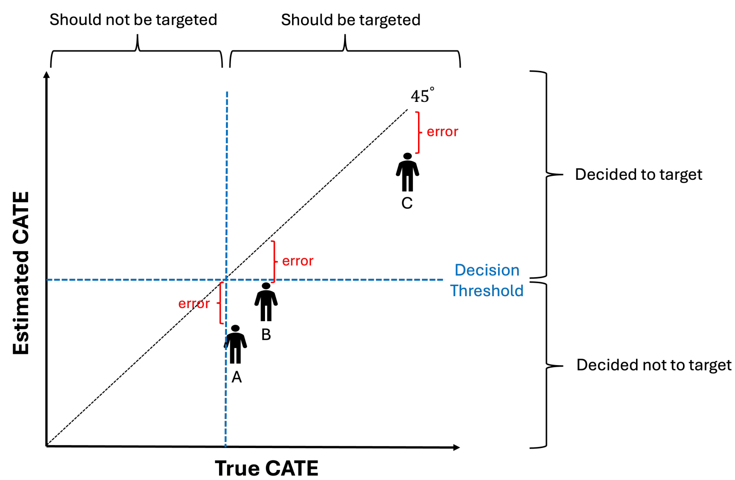


Figure 1: Differential Impact of Prediction Errors on Targeting Profitability
Customers A, B, and C have identical prediction errors, but their profit implications differ: errors matter most when true CATE is near the threshold and profitability is high.

Conversely, for customers A and B, whose incremental spending is close to $2 (i.e., whose true CATEs are near the decision threshold), even small prediction errors can lead to mistargeting. Importantly, the impact of mistargeting these two customers differs in terms of profitability. Mistargeting customer A has minimal effect because the profit earned by the company from customer A remains nearly unchanged regardless of whether they are targeted. By contrast, mistargeting customer B leads to a notable profit loss, as their true incremental spending (CATE) is actually greater than $2.

In general, Figure 1 highlights a key insight: when firms rely on CATE estimates to deploy targeted policies, not all customers contribute equally to decision quality. In particular, greater attention should be directed toward customers for whom prediction errors are most consequential for profitability. Intuitively, these are the customers whose true CATEs lie near — but do not precisely coincide with — the decision threshold.

2.1. Formal Setup

To formalize this problem, we consider a setting where each customer i is described by a vector of pre-treatment covariates $\mathbf{X}_i$ and can be assigned to a binary treatment $W_i \in \{0, 1\}$, with $W_i = 1$ indicating exposure to the marketing intervention and $W_i = 0$ representing the control condition. The potential outcomes under treatment and control are $Y_i(1)$ and $Y_i(0)$, respectively, and the observed outcome is $Y_i = Y_i(W_i)$.

Firm's objective The firm seeks to develop a targeting policy that, based on the observed covariates $\mathbf{X}_i$, assigns each customer to treatment or control in order to maximize expected profitability. This objective leads to an optimal decision rule $\pi^*(\cdot) : \mathbf{X} \to \{0, 1\}$ defined as:

$$\begin{aligned}
\pi^*(\mathbf{X}_i) &= \arg\max_{\pi} \mathbf{E}_{\mathbf{X}}[\underbrace{Y_i(0) \cdot (1 - \pi(\mathbf{X}_i))}_{\text{profit w/o treatment}} + \underbrace{(Y_i(1) - c(\mathbf{X}_i)) \cdot \pi(\mathbf{X}_i)}_{\text{profit w/ treatment}} |\mathbf{X}_i] \\
&= \arg\max_{\pi} \mathbf{E}_{\mathbf{X}}[Y_i(0)|\mathbf{X}_i] + (\underbrace{\mathbf{E}[(Y_i(1) - Y_i(0))|\mathbf{X}_i]}_{\text{CATE}} - c(\mathbf{X}_i)) \cdot \pi(\mathbf{X}_i) \\
&= \arg\max_{\pi} \mathbf{E}_{\mathbf{X}}[Y_i(0)|\mathbf{X}_i] + \underbrace{(\tau(\mathbf{X}_i) - c(\mathbf{X}_i))}_{\text{incremental profit}} \cdot \pi(\mathbf{X}_i),
\end{aligned} \tag{1}$$

where

$$\tau(\mathbf{X}_i) = \mathbf{E}_{\mathbf{X}}[Y_i(1) - Y_i(0)|\mathbf{X}_i]$$

is the CATE, or the expected incremental impact of the intervention conditional on covariates $\mathbf{X}_i$.[1]

[1] The expectation here is taken with respect to the covariate distribution of the firm's customer base.

As Equation (1) shows, the firm's decision rule depends solely on the incremental profit from the intervention; that is, the difference between the CATE and the cost of treatment. Therefore, if the firm had access to the true CATE, the optimal targeting decision would be:

$$\pi^*(\mathbf{X}_i) = \mathbf{1}\{\tau(\mathbf{X}_i) > c(\mathbf{X}_i)\}. \tag{2}$$

Firm's inference problem In practice, however, the true CATE is never observed, as for each customer we see only one potential outcome (Holland 1986). Instead, firms must infer it from data. The standard approach involves conducting a randomized experiment in which a subset of customers S sampled from the firm's customer base $\mathcal{I}$ is randomly assigned to treatment or control. Using the outcomes from this experimental sample, the firm estimates a CATE model, which enables them to predict the CATE for any new customer i, $\hat{\tau}_S(\mathbf{X}_i)$ and construct a policy:

$$\hat{\pi}_S(\mathbf{X}_i) = \mathbf{1}\{\hat{\tau}_S(\mathbf{X}_i) > c(\mathbf{X}_i)\}. \tag{3}$$

This formalization highlights several points. First, targeting decisions are based on *estimated* CATEs from the experimental sample S, where prediction errors are inevitable due to learning from finite samples. As a result, the learned policy $\hat{\pi}_S$ will coincide with the optimal policy π^* only when estimation is sufficiently accurate. Second, since the policy π is implemented via a threshold rule — treating only when the estimated CATE exceeds the intervention cost — the consequences of prediction errors in $\hat{\tau}_S(\mathbf{X}_i)$ are non-linear and vary across customers. Third, the profitability impact of these errors depends not only on whether a customer is mistargeted but also on how profitable that customer would have been if treated correctly (i.e., their opportunity cost). Mistargeting high-value customers leads to disproportionately large losses in potential profits.

Taken together, the firm's ability to make profitable targeting decisions hinges on the composition of the experimental sample S, which in turn affects the type and magnitude of prediction errors in $\hat{\tau}_S(\mathbf{X}_i)$. This dependence becomes particularly consequential when the experimental sample size is limited and estimation uncertainty is high, as is often the case in marketing settings. In such cases, *who* the firm samples can play a critical role in the profitability of the resulting targeting policy.

Policy-Aware Experimentation This leads to the central question we address: how should firms collect experimental samples in a way that improves the profitability of the resulting targeting policy?

We formalize that objective as, given a fixed number of customers N_S to be part of the experimental sample S, the firm aims to choose a sample S^* such that the targeting policy $\hat{\pi}_S$ estimated from it maximizes the firm's expected profit when deployed on future customers with the same distribution as $\mathcal{I}$:[2]

$$S^* = \max_{S \in \mathcal{I}, |S| = N_S} f(S) \tag{4}$$
$$f(S) = \mathbf{E}\left[Y_i(0) + (\tau(\mathbf{X}_i) - c(\mathbf{X}_i)) \cdot \hat{\pi}_S(\mathbf{X}_i)\right],$$

where the expectation is taken with respect to both the covariate distribution of firm's customer base and the sampling distribution of $\hat{\pi}_S$.

Framing the sampling problem as a pure combinatorial optimization — selecting the most informative subset of customers from a large population — quickly becomes infeasible, as evaluating all $\binom{|\mathcal{I}|}{N_S}$ possible combinations is computationally infeasible in practice. This challenge calls for a more principled and scalable solution. Building on the economic intuition developed above, we introduce a tractable and theoretically grounded approach that prioritizes customers whose treatment effect estimation errors are most likely to influence targeting profitability. Specifically, we propose a sequential experimental design guided by a novel sampling criterion, which we term *expected profit loss (EPL)*. By directing experimental resources toward *consequential customers*, EPL enables firms to strategically align data collection with their profit-maximization objectives.

Building on this foundation, the remainder of the paper introduces and evaluates our policy-aware experimentation framework. We begin with a review of related work. Section 4 formalizes the expected profit loss (EPL) criterion and presents our proposed sequential sampling algorithm. Section 5 uses simulation studies to benchmark performance and illustrate the economic gains from our approach. Section 6 offers empirical validation through two empirical applications in customer reactivation and mobile promotions. Finally, Section 7 concludes with a discussion of practical implications and avenues for future research.

[2] This assumes no concept drift (stable treatment effects) and no covariate shift (stable customer characteristics distribution) between the experimental period and deployment period, consistent with standard literature assumptions (e.g. Simester et al. 2020; Huang et al. 2024).

3. Related Literature

This paper relates and connects four main streams of literature: targeting and personalization, experimental design, active learning, and decision-aware learning.

Targeting and personalization. A large body of research has explored how firms can use customer-level data to personalize marketing interventions (e.g., Ascarza 2018; Simester et al. 2020; Ellickson et al. 2022; Yang et al. 2023; Huang and Ascarza 2024; Hitsch et al. 2024). Recent work has emphasized the importance of estimating heterogeneous treatment effects and developing policy learning methods that directly map customer characteristics to targeting decisions (Zhao et al. 2012; Athey and Wager 2021, e.g.). These approaches shift focus from estimating conditional average treatment effects (CATEs) to learning decision rules that maximize business outcomes. However, while these methods align the *inference stage* with the firm's objective, they typically rely on experimental data collected using generic sampling strategies. To the best of our knowledge, prior work has not incorporated the firm's business objective into the *experimental design* stage itself, leaving a critical misalignment between how data are collected and how they are ultimately used.

Experimental design. Our work relates to the broader experimental design literature, which we divide into two main streams. The first focuses on minimizing estimation uncertainty, typically through classical parametric designs such as A- and D-optimality (e.g., Kiefer and Wolfowitz 1959; Fontaine et al. 2020), or through optimal experimental sizing and allocation across customer segments (e.g., Feit and Berman 2019; Simester et al. 2022; Hu et al. 2024). The latter includes the literature on adaptive experiments, which draws from multi-armed bandits (e.g., Schwartz et al. 2017; Misra et al. 2019; Waisman et al. 2024) and best arm identification (BAI) methods (e.g., Bubeck et al. 2010; Kasy and Sautmann 2021; Kato et al. 2024). While our sequential design is consistent with this literature, our approach shifts the focus from adaptively assigning treatment conditions to selecting *whom* to sample in the first place, in order to improve downstream decision quality. Moreover, we show that our method remains effective even in a simple two-stage design, providing a more practical and scalable alternative to fully adaptive frameworks.

Active learning. Third, our research is closely related to the active learning literature, which develops acquisition strategies to reduce the cost of data collection (e.g., Fu et al. 2013; Shankaranarayana 2023). While most of this work focuses on improving predictive accuracy for supervised learning tasks, more recent efforts have turned to treatment effect estimation (e.g., Puha et al. 2020; Jesson et al. 2022) and decision-making tasks (e.g., Sundin et al. 2019; Liu et al. 2023). Our approach builds on this stream by proposing a new acquisition function — expected profit loss (EPL) — that prioritizes customers whose prediction errors are most consequential for targeting profitability. In doing so, we connect active learning with firm-level objectives in a policy-driven context.

Decision-aware learning. Finally, this paper contributes to the literature on decision-aware learning, which seeks to integrate prediction with downstream optimization in the context of the predict-then-optimize framework (e.g., Wilder et al. 2019; Kallus and Mao 2023). While much of this literature focuses on refining predictive models using observed outcomes to improve decision quality (e.g., Lemmens and Gupta 2020; Chung et al. 2022), our work makes two important advancements. First, we move attention to an earlier stage: the design of the experiment itself. Second, we address the more challenging context of counterfactual reasoning, where the outcomes of interest — CATEs — are unobserved. In doing so, we offer a novel perspective on aligning experimental design with business objectives in marketing.

In sum, our approach integrates ideas from active learning, adaptive experimentation, and decision-aware learning to create a unified framework that aligns the firm's business objective with the entire empirical strategy — from how data are collected to how targeting policies are ultimately deployed. By embedding profit considerations into the sampling stage, we offer a decision-aware perspective that enhances both the efficiency and effectiveness of marketing experimentation for policy learning.

4. Methodology

This section introduces our methodology for selecting an experimental sample, S, that enhances policy quality. In particular, we introduce the expected profit loss criterion (EPL), which quantifies how prediction errors in CATE estimation impact firm's profitability. We provide theoretical guarantees for using this criterion as the basis for the firm's sampling strategy, establishing a strong foundation for its use in policy-aware experimentation.

To implement EPL sampling in practice, we develop two complementary strategies. First, we develop an *estimation approach for EPL* that leverages Bayesian inference for uncertainty quantification within a Causal Forest framework (Athey et al. 2019b), providing a principled and scalable method to identify high-impact customers without prior knowledge of treatment effects. Second, we propose a *sequential experimental design* that begins with uniform sampling and then adaptively reallocates experimental resources toward individuals with the highest EPL estimates. This dynamic sampling strategy concentrates learning on the most economically relevant segments of the population.

We conclude the section by discussing key implementation considerations and the practical advantages of our proposed design.

4.1. Near-Optimal Sampling Strategy: The Expected Profit Loss (EPL) Sampling Approach

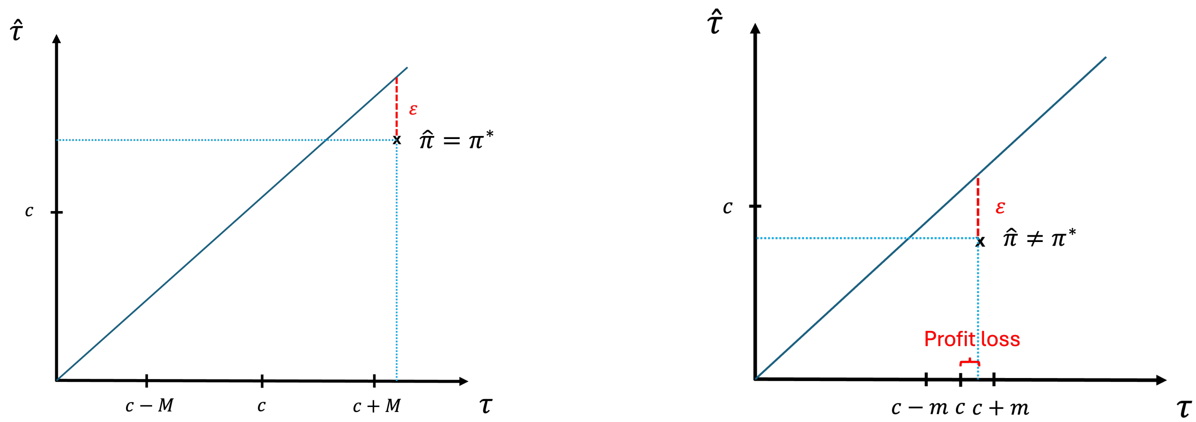
We begin by formally characterizing *consequential customers* — whose prediction errors are most harmful for profitability and thus merit greater attention during experimentation. This characterization is the basis for our expected profit loss (EPL) formulation.

4.1.1. Consequential Customers and Expected Profit Loss

Building on the intuition from Figure 1, consider the two illustrative scenarios depicted in Figure 2. In the left panel, the customer's true CATE exceeds the intervention cost by a large margin, i.e., $\tau(\mathbf{X}_i) - c(\mathbf{X}_i) > M$. In this case, even a moderate prediction error ε is unlikely to alter the targeting decision. By contrast, the right panel depicts a customer whose true CATE lies close to the threshold, i.e., $\tau(\mathbf{X}_i) - c(\mathbf{X}_i) > m$, where a small estimation error can easily lead to mistargeting.

Importantly, the probability of mistargeting alone does not fully capture a customer's contribution to decision quality. The associated profit loss incurred when such mistargeting occurs (as highlighted in Figure 2b) is equally critical. Customers whose CATEs are nearly equal to the intervention cost (i.e., $\tau(\mathbf{X}_i) \approx c(\mathbf{X}_i)$) yield minimal profit gain or loss regardless of whether they are treated. In contrast, the most meaningful profit losses arise from customers whose CATE deviates substantially from the threshold but are still misclassified due to estimation error.

Therefore, we characterize *consequential customers* as individuals who satisfy two criteria: (1) a high probability of mistargeting, $\mathbb{P}\left(\hat{\pi}(\mathbf{X}_i) \neq \pi^*(\mathbf{X}_i)\right) = \mathbb{E}\left[\mathbf{1}\{\hat{\pi}(\mathbf{X}_i) \neq \pi^*(\mathbf{X}_i)\}\right]$, and (ii) a non-



(a) $\tau(\mathbf{X}_i) - c(\mathbf{X}_i) > M$ **(b)** $\tau(\mathbf{X}_i) - c(\mathbf{X}_i) > m$

Figure 2: Differential Impact of Prediction Error ε on Targeting Accuracy

trivial absolute profit loss: $|\tau(\mathbf{X}_i) - c(\mathbf{X}_i)|$, and formally define *expected profit loss (EPL)* as:

$$\ell(\mathbf{X}_i) = \mathbb{E}_{\hat{\pi}}\left[\underbrace{|\tau(\mathbf{X}_i) - c(\mathbf{X}_i)|}_{\text{profit loss}} \cdot \underbrace{\mathbf{1}\{\hat{\pi}(\mathbf{X}_i) \neq \pi^*(\mathbf{X}_i)\}}_{\text{mistargeting prob.}}\right],$$

where the expectation is taken with respect to the sampling distribution of $\hat{\pi}$. This quantity should guide the firm to sample consequential customers for policy-aware experimentation.

To operationalize this approach, firms must rely on observed data to estimate $\ell(\mathbf{X}_i)$. This requires an initial experimental sample, which we denote as $\mathcal{D}$. In the next section, we provide theoretical guarantees for the EPL-based sampling strategy, assuming the existence of an initial sample. In Section 4.2 we then outline how firms can construct this initial sample, estimate the EPL function in practice through sequential sampling, and ultimately prioritize experimental resources toward consequential customers.

4.1.2. Theoretical Guarantee of EPL Sampling

The following proposition shows that allocating experimental resources based on EPL yields a strong approximation to the optimal sampling strategy such that the targeting policy estimated from it maximizes the firm's expected profit when deployed on future customers.

Suppose the firm has collected an initial experimental dataset $\mathcal{D}$ by randomly sampling customers and assigning treatments. This dataset enables estimation of CATEs and EPLs, which guides subsequent sampling decisions.

Proposition 1 (Near-Optimality of Expected Profit Loss Sampling). *Let $\mathcal{I}$ denote the set of customers available for experimentation. Let N_S denote the desired total sample size (with $N_S > |\mathcal{D}|$). Under suitable regularity conditions, for any sufficiently large initial sample $\mathcal{D}$, a greedy algorithm that iteratively selects $k = N_S - |\mathcal{D}|$ additional customers with the highest marginal expected profit loss,*

$$\ell_S(\mathbf{X}_i) = \mathbb{E}_{\hat{\pi}}[|\tau(\mathbf{X}_i) - c(\mathbf{X}_i)| \cdot \mathbf{1}\{\hat{\pi}_S(\mathbf{X}_i) \neq \pi^*(\mathbf{X}_i)\}] \tag{5}$$

achieves a $(1 - \frac{1}{e})$-approximation to the globally optimal sampling strategy, i.e.,

$$f(\mathcal{D} \cup S_k^g) \geq \left(1 - \frac{1}{e}\right) f(\mathcal{D} \cup S_k^*)$$

where S_k^g denotes the k customers selected by the greedy algorithm and S_k^* denotes the optimal set of k customers such that the targeting policy $\hat{\pi}_S$ estimated from it maximizes the firm's expected profit when deployed on future customers with the same distribution as $\mathcal{I}$. That is,

$$S_k^* = \arg \max_{S_k \subseteq \mathcal{I}, |S_k| = k} \mathbb{E}\left[Y_i(0) + (\tau(\mathbf{X}_i) - c(\mathbf{X}_i)) \cdot \hat{\pi}_{\mathcal{D} \cup S_k}(\mathbf{X}_i)\right], \tag{6}$$

where the expectation is taken with respect to both the covariate distribution of firm's customer base and the sampling distribution of $\hat{\pi}_S$.

Proof. See Web Appendix B. □

Proposition 1 demonstrates that the expected profit loss (EPL) sampling strategy achieves near-optimality by attaining a $(1 - \frac{1}{e})$-approximation to the globally optimal sampling policy defined in Equation (4). This result confirms that selecting customers with the highest EPLs aligns experimental design with the firm's profit-maximizing objective, offering theoretical performance guarantees of our proposed approach. Note that the requirement for a sufficiently large initial sample set D in Proposition 1 is not a constraint in practice, as such a sample is already needed to estimate EPLs — an implementation step we describe next in Section 4.2. This initial data collection simultaneously satisfies the theoretical condition and enables estimation of EPLs.

While Proposition 1 establishes strong guarantees for EPL-based sampling, applying this strategy in practice requires estimating EPLs without observing true CATEs — a challenge we address in the next section.

4.2. Practical Implementation: A Sequential Experimental Design with EPL Sampling

While the EPL sampling strategy provides a theoretically grounded framework for optimizing firm's profitability, its implementation entails a major challenge due to the fundamental problem of causal inference. Specifically, the computation of $\ell(\mathbf{X}_i)$ depends on the true CATE, $\tau(\mathbf{X}_i)$, and the corresponding optimal decision $\pi^*(\mathbf{X}_i)$, both of which are unknown to the firm. As such, $\ell(\mathbf{X}_i)$ cannot be used directly as a sampling criterion in practice. Instead, firms must estimate an approximation, which we denote $\hat{\ell}(\mathbf{X}_i)$, to guide the sampling process.

In this section, we tackle this challenge by developing an estimation strategy that enables accurate identification of customers with the highest EPL, even without direct observation of true CATEs. As discussed earlier this requires an initial experimental sample D that is sufficiently large and representative, which is rarely the case. To address this challenge, we adopt a sequential design that begins with a small pilot sample and progressively refines EPL estimates as new data are collected. This structure enables the algorithm to improve its prioritization of consequential customers over time, offering a flexible and scalable path toward approximating the theoretical ideal. Together, these components provide a practical and scalable implementation of our theoretically founded sampling strategy.

4.2.1. Estimating Expected Profit Loss (EPL)

We present our estimation strategy for EPL, which enables firms to identify consequential customers without directly observing the true CATEs. Our approach leverages Bayesian inference for uncertainty quantification, operationalized through the Causal Forest model (Wager and Athey 2018; Athey et al. 2019b) to avoid restrictive parametric assumptions. Specifically, we use the ensemble of trees generated by the Causal Forest to approximate the posterior predictive distribution of the CATE, following the well-established result that bootstrapped trees approximate a posterior under a noninformative prior (Hastie et al. 2009).[3] Each tree, trained on a different bootstrapped subsample, yields a distinct prediction of the treatment effect for a given customer. We treat these tree-level predictions as samples from the posterior distribution, forming an empirical approximation that captures model uncertainty. This uncertainty-aware view of CATE estimation allows

[3]We chose Causal Forest due to its practical implementation and its local property required in our theoretical analysis. Firms can implement EPL estimation using `grf` in R or `econML` in Python.

us to construct an EPL score that quantifies how much profit a firm stands to lose from potentially mistargeting a customer.

We implement this approach by first using the previously collected dataset $\mathcal{D}$ to train a Causal Forest model. For each customer i outside the training set, we compute the posterior mean of the CATE, $\hat{\tau}_{\mathcal{D}}(\mathbf{X}_i)$, and treat the predictions from each tree j in the ensemble as draws $\tilde{\tau}^j_{\mathcal{D}}(\mathbf{X}_i)$, forming an empirical approximation of the posterior predictive distribution $p_{\mathcal{D}}(\tilde{\tau}(\mathbf{X}_i) \mid \mathcal{D})$. This distribution enables us to assess the likelihood that the estimated targeting decision, $\hat{\pi}_{\mathcal{D}}(\mathbf{X}_i) = \mathbf{1}\{\hat{\tau}_{\mathcal{D}}(\mathbf{X}_i) > c(\mathbf{X}_i)\}$, deviates from the optimal decision implied by each draw.

We then define the EPL for customer i as the expected profit loss from potential deviations between the estimated and optimal targeting decisions, integrated over the posterior predictive distribution:

$$
\begin{aligned}
\hat{\ell}_{\mathcal{D}}(\mathbf{X}_i) &= \int \underbrace{0 \cdot \mathbf{1}\{\hat{\pi}_{\mathcal{D}}(\mathbf{X}_i) = \tilde{\pi}(\mathbf{X}_i)\}}_{\text{profit loss with no deviation}} + \underbrace{|\tilde{\pi}(\mathbf{X}_i) - c(\mathbf{X}_i)|\mathbf{1}\{\hat{\pi}_{\mathcal{D}}(\mathbf{X}_i) \neq \tilde{\pi}(\mathbf{X}_i)\}}_{\text{profit loss with deviation}} p_{\mathcal{D}}(\tilde{\tau}(\mathbf{X}_i)|\mathcal{D})d\tilde{\tau} \\
&= \int \underbrace{|\tilde{\tau}(\mathbf{X}_i) - c(\mathbf{X}_i)|\mathbf{1}\{\hat{\pi}_{\mathcal{D}}(\mathbf{X}_i) \neq \tilde{\pi}(\mathbf{X}_i)\}}_{\text{profit loss}} p_{\mathcal{D}}(\tilde{\tau}(\mathbf{X}_i)|\mathcal{D})d\tilde{\tau},
\end{aligned}
\tag{7}
$$

where $\tilde{\tau}(\mathbf{X}_i)$ is a draw from the posterior predictive distribution and $\tilde{\pi}(\mathbf{X}_i) = \mathbf{1}\{\tilde{\tau}(\mathbf{X}_i) > c(\mathbf{X}_i)\}$.

For each draw, we determine whether the implied targeting decision differs from the estimated one and the corresponding profit loss $|\tilde{\tau}^j_{\mathcal{D}}(\mathbf{X}_i) - c(\mathbf{X}_i)|\mathbf{1}\{\hat{\pi}_{\mathcal{D}}(\mathbf{X}_i) \neq \tilde{\pi}^j_{\mathcal{D}}(\mathbf{X}_i)\}$ for each tree. We then approximate the integral in Equation (7) by averaging the profit losses $\tilde{\tau}^j_{\mathcal{D}}(\mathbf{X}_i)$ across all J trees to obtain the EPL estimate, $\hat{\ell}_{\mathcal{D}}(\mathbf{X}_i)$. The pseudo-code for the estimation algorithm is outlined in Algorithm 1.

Intuitively, our estimated EPL closely approximates the true EPL ranking. When the true CATE $\tau(\mathbf{X}_i)$ is far from the decision threshold, the posterior mean $\hat{\tau}_{\mathcal{D}}(\mathbf{X}_i)$ and most draws from the posterior predictive distribution are likely to fall on the same side of the threshold. Thus, few draws produce a targeting decision different from $\hat{\pi}_{\mathcal{D}}(\mathbf{X}_i)$, yielding low estimated EPL that correctly reflects low true EPL. When the true CATE is very close to the threshold, the posterior distribution concentrates around the threshold, making the profit loss $|\tilde{\tau}(\mathbf{X}_i) - c(\mathbf{X}_i)|$ small for most draws and producing modest estimated EPL that again aligns with modest true EPL. The highest estimated EPL values arise when the true CATE is moderately distant from the threshold—creating substantial potential profit loss—while posterior uncertainty generates many draws that cross the threshold and produce opposite targeting decisions. These customers have both high true and estimated EPL, ensuring our algorithm prioritizes the right customers.

Algorithm 1 Estimating Expected Profit Loss

Input: Experimental data $\mathcal{D}$; Number of trees $J \in \mathbb{N}$
Output: $\hat{\ell}_{\mathcal{D}}(\mathbf{X}_i)$
Data: $\mathcal{D}$
Train a Causal Forest model on $\mathcal{D}$ and obtain $\hat{\pi}_{\mathcal{D}}(\mathbf{X}_i) = \mathbf{1}\{\hat{\tau}_{\mathcal{D}}(\mathbf{X}_i) > c(\mathbf{X}_i)\}$.
for j = 1, 2, ..., J **do**
Determine the optimal targeting decision associated with the prediction $\tilde{\tau}^j_{\mathcal{D}}(\mathbf{X}_i)$:

$$\tilde{\pi}^j_{\mathcal{D}}(\mathbf{X}_i) = \mathbf{1}\{\tilde{\tau}^j_{\mathcal{D}}(\mathbf{X}_i) > c(\mathbf{X}_i)\}.$$

Calculate the profit loss associated with the prediction $\tilde{\tau}^j_{\mathcal{D}}(\mathbf{X}_i)$ by

$$\hat{\ell}^j_{\mathcal{D}}(\mathbf{X}_i) = |\tilde{\tau}^j_{\mathcal{D}}(\mathbf{X}_i) - c(\mathbf{X}_i)|\mathbf{1}\{\hat{\pi}_{\mathcal{D}}(\mathbf{X}_i) \neq \tilde{\pi}^j_{\mathcal{D}}(\mathbf{X}_i)\}.$$

end for
Calculate the EPL estimate by averaging the profit losses associated with the draws across all J trees:

$$\hat{\ell}_{\mathcal{D}}(\mathbf{X}_i) = \frac{1}{J}\sum_{j=1}^{J} \hat{\ell}^j_{\mathcal{D}}(\mathbf{X}_i).$$

Return: EPL estimate $\hat{\ell}_{\mathcal{D}}(\mathbf{X}_i)$

4.2.2. A Sequential Experimental Design with Expected Profit Loss Sampling

With an estimation procedure for EPL in place, we next describe how these scores can be operationalized to guide sequential experimentation. As described in Section 4.2.1, EPL estimation requires experimental data, creating a circular dependency between identifying consequential customers and collecting the data needed to do so. To address this challenge, we propose a multi-stage sequential design that begins with an initial randomized batch of customers — forming a dataset we previously denoted $\mathcal{D}$ — used to train a preliminary CATE model. The resulting posterior from this model enables the estimation of EPL scores, which in turn guides the prioritization of customers in subsequent experimental batches.

Importantly, as more data is collected, the accuracy of EPL estimates improves, allowing the procedure to increasingly focus on customers whose mistargeting would lead to substantial profit loss. While a two-stage design can suffice, a multi-stage sequential approach can offer superior performance by allowing for more accurate sampling of consequential customers and progressive refinement of customer prioritization. In particular, it enables a more effective trade-off between gathering information to improve EPL estimates and focusing resources on customers with high estimated EPL. We formalize the multi-stage sequential procedure below and treat the two-stage design as a simplified special case.

Sequential procedure Given a customer base $\mathcal{I}$ and a pre-determined experimental size N_S (with $N_S < |\mathcal{I}|$), we obtain the experiment sample S by combining B sequential batches $S = \cup_{b=1}^{B} S_b$. Each batch $b \in \{1, 2, ..., B\}$ contains $|S_b| = n_b$ customers such that $\sum_{b=1}^{B} n_b = N_S$.[4] In the first batch ($b = 1$), we randomly sample $|S_1| = n_1$ customers from $\mathcal{I}$ and assign them randomly to the two treatment arms $W_i \in \{0, 1\}$. After observing their outcomes, we use this data S_1 to estimate EPLs for the remaining customers.[5] In each subsequent batch $b = 2, \ldots, B$, we select $|S_b| = n_b$ customers with the highest estimated EPLs among those not yet sampled, i.e., from $\mathcal{I} \setminus S^{b-1}$ where $S^{b-1} = \cup_{j=1}^{b-1} S_j$, and again assign treatments at random. This process continues until the final batch B. Once outcomes are observed from all B batches, we re-estimate a CATE model on the full experimental dataset $S = \cup_{b=1}^{B} S_b$.

Crucially, our design differs from commonly used bandit or best-arm identification strategies (e.g., Thompson Sampling (Schwartz et al. 2017; Jain et al. 2024) or UCB (Misra et al. 2019)) that adaptively assign treatments. Here, treatment assignment W is always random; what differs with respect to standard practice is *which customers are prioritized* for inclusion in the experiment.

Pseudocode for the full procedure is provided in Algorithm 2.

Assumptions In line with prior work, we impose the following assumptions across waves of the experiment:

Assumption 1. *(Stable Unit Treatment Value Assumption, SUTVA) The potential outcomes for customer i depend solely on their own treatment assignment, not on the assignment of any other customer i'. Formally,*

$$Y_i(\mathbf{W}) = Y_i(W_i).$$

[4] We focus on scenarios where the firm has access to a fixed customer base, excluding cases with streaming arrivals. This assumption reflects many marketing applications such as retention (e.g. Ascarza 2018; Lemmens and Gupta 2020; Yang et al. 2023) and catalog mailings (e.g. Hitsch et al. 2024; Simester et al. 2020).

[5] This initial batch S_1 corresponds to the dataset $\mathcal{D}$ introduced in Section 4.2.1.

Algorithm 2 The Sequential Experimental Design with Expected Profit Loss Sampling

Customer Base: $\mathcal{I}$
Input: Experimental size N_S; Number of batches B; Number of samples allocated to each batch $\{n_b\}_{b=1}^B$ with $\sum_{b=1}^B n_b = N_S$
Output: Targeting Decision $\hat{\pi}_S$
for b = 1, 2, ..., B **do**
  if b=1 **then**
    Randomly sample n_1 customers from $\mathcal{I}$
    Randomly assign each sampled customer i to the two treatment arms $W_i \in \{0, 1\}$
    Observe the outcomes of the sampled customers Y_i
  else
    For each unsampled customer $i \notin S^{b-1} = \cup_{j=1}^{b-1} S_j$, estimate the EPL $\hat{\ell}_{S^{b-1}}(\mathbf{X}_i)$ with the data collected in previous $b-1$ batches S^{b-1} (Algorithm 1)
    Select n_b customers with the highest EPL estimates $\hat{\ell}_{S^{b-1}}(\mathbf{X}_i)$
    Randomly assign each customer i in this batch to the two treatment arms $W_i \in \{0, 1\}$
    Observe the outcomes of the sampled customers Y_i
  end if
end for
Estimate a CATE model with the experimental data S
Derive the final targeting decision leveraging the CATE predictions:

$$\hat{\pi}_{S^B}(x) = \mathbf{1}\{\hat{\tau}_{S^B}(x) > c(x)\}.$$

Return: Targeting decision $\hat{\pi}_{S^B}$

Assumption 2. *(Stability) The distribution of potential outcomes for any customer i is invariant across time. That is, for all $t \neq t'$,*

$$\mathbb{E}_{\mathbf{X}}[Y_i^t(W_i)|\mathbf{X}_i] = \mathbb{E}_{\mathbf{X}}[Y_i^{t'}(W_i)|\mathbf{X}_i],$$

where $Y_i^t(W_i)$ and $Y_i^{t'}(W_i)$ denote the potential outcomes at times t and t', respectively.

Assumption 3. *(Unconfoundedness) For each customer i, treatment assignment is independent of unobserved potential outcomes, conditional on covariates:*

$$(Y_i(0), Y_i(1)) \perp\!\!\!\perp W_i \mid \mathbf{X}_i.$$

Assumption 4. *(Overlap) Each customer has a non-zero probability of receiving either treatment condition:*

$$0 < \Pr(W_i = 1 \mid \mathbf{X}_i = x) < 1, \quad \forall x.$$

Assumption 1 rules out spillover effects across customers, while Assumption 2 ensures that treatment effects remain stable over the experimental horizon. Together, these conditions imply that customers identified as high-priority targets early on will remain so in subsequent waves, facilitating efficient adaptive allocation.

Assumptions 3 and 4 are automatically satisfied given that each customer in the experiment is randomly assigned to one of the two treatment arms, which guarantees that each experimental wave yields consistent CATE estimates for $\tau(\mathbf{X}_i)$.[6] These estimates form the basis for the final targeting policy outlined in Section 2.

Summary Our sequential procedure begins with a pilot experiment of size n_1, which is used to train a flexible model of heterogeneous treatment effects. This initial step mirrors the conventional "test-then-learn" paradigm (e.g., Yoganarasimhan et al. 2023; Huang and Ascarza 2024), with the key distinction that only the pilot stage employs uniform sampling. At each subsequent wave $b = 2, \ldots, B$, the firm (i) computes the EPL estimates for each customer based on data collected in previous batches; (ii) selects n_b customers with the highest estimated EPLs (i.e., those associated with the highest stepwise regret); (iii) assigns the selected customers to treatment or control randomly and observes their outcomes; and (iv) updates the CATE model using the newly accrued data before proceeding to the next wave. Because each wave prioritizes individuals for whom errors in $\tau(X_i)$ estimation are most economically consequential, the design adaptively allocates experimental resources to customers with the greatest value of information.

4.3. Implementation Considerations

Implementing our proposed sequential experimental design involves two key design decisions: (1) determining the number of batches in the experiment, and (2) allocating samples across these batches. In this section, we offer practical guidance to help firms navigate these design choices, drawing on extensive simulation studies and empirical analyses that inform the robustness and effectiveness of our recommendations.

Number of batches. Firms face a trade-off between targeting precision and operational feasibility when choosing the number of batches, B. Increasing B enables more frequent updates of EPL estimates and sharper identification of consequential customers. However, it also extends the experimental timeline and introduces operational complexity. We recommend setting B to

[6] Note that under this design, the sample $\mathcal{S}$ may not be representative of the population, and hence estimates of the average treatment effect (ATE) may be biased. However, standard reweighting techniques can be applied to recover unbiased ATE estimates from non-representative samples.

the maximum number feasible given the firm's cadence of outcome measurement. For example, in settings where outcomes are observed quickly (e.g., digital engagement), a design with 3–5 batches may be practical. In contrast, when outcomes require longer observation windows (e.g., revenue after a promotion), a two-stage design may be more suitable. To add flexibility, firms can adopt an *early stopping rule* — terminating the experiment once EPL estimates stabilize or fall below a pre-specified threshold.

Batch size allocation. Once B is set, the firm must allocate its experimental budget across batches. This introduces another trade-off: allocating more samples to earlier batches improves EPL estimation, while saving samples for later batches preserves flexibility for targeting high-EPL customers. A *decreasing batch size configuration* — in which earlier batches receive more samples — can improve estimation accuracy without sacrificing performance. However, our simulation (Appendix C.2.1) and empirical results (Appendix D.4.1) suggest that both constant and decreasing allocations perform similarly in practice. We encourage firms to choose the scheme that aligns with their operational rhythm and technical infrastructure while ensuring adequate initial sample size.

When to use two-stage vs. multi-stage designs. A key practical challenge in adaptive experimentation is the time lag between treatment and observable outcomes. This becomes particularly salient when the desired outcomes (e.g., purchases, revenue) take days or weeks to manifest.[7] Delays can make a fully adaptive design slow or operationally burdensome, especially when combined with engineering costs related to frequent model updates (Hadad et al. 2021). In such settings, a simplified two-stage version of our approach may be more suitable.

In a two-stage design, the experiment is split into two (potentially unequal) groups. The first-stage sample is used to estimate EPLs, and the second-stage sample is drawn based on these estimates. This simplification reduces complexity and shortens timelines, while retaining most of the performance benefits of a full multi-stage design (see Sections 5 and 6).

The two-stage design introduces a single trade-off: the first-stage sample must be large enough to generate accurate EPL estimates, while leaving enough customers in the second stage to act on those estimates. We explore this trade-off empirically in Appendix C.2.2 and D.4.2.

[7]One may argue that this can be addressed by using intermediate outcomes as surrogates for delayed feedback (Athey et al. 2019a; Yang et al. 2023; Huang and Ascarza 2024). However, these proxies may themselves require substantial time to observe post-intervention.

5. Simulation

Before testing our approach on real data, we perform a series of simulation studies to achieve two main objectives. First, we examine the performance of our method in learning the profit-maximizing targeting policy across various scenarios, identifying conditions under which our approach provides the greatest benefits. We evaluate both the multi-stage and the two-stage design, and find that the latter maintaining strong performance while offering a more practical alternative to fully adaptive design, as discussed in Section 4.3. Second, we leverage the advantages of synthetic data where true CATE is known to evaluate the ability of our approach to accurately identify customers whose prediction errors in CATE estimation have the greatest impact onprofitability, as illustrated in Section 4.1.1.

5.1. Simulation Setup

We consider the scenario where a firm aims to develop a targeting policy that maximizes the profitability of a marketing intervention. The firm first learns the targeting policy through experimentation on a subset of customers, and subsequently implements the learned policy on future customers. We assume that the impact of the (binary) intervention on customers is heterogeneous and follows a normal distribution centered at 1, meaning on average, the intervention has a positive effect on consumers.[8]

Specifically, we generate a customer base $\mathcal{I}$ and an evaluation set D_{eval} with a binary treatment $W_i \in \{0, 1\}$ according to the following data generating process:

$$Y_i = \tau(X_i) \cdot W_i + X_{i4} \cdot X_{i5} + e_i$$

where

$$\begin{aligned}
\tau(X_i) &= X_{i1} \cdot X_{i2} + 0.5 \cdot X_{i3} + 1, \\
e_i &\sim \mathcal{N}(0, 1), \\
X_{ij} &\sim \mathcal{N}(0, 1), \quad j \in \{1, 3, 5\}, \\
X_{ij} &\sim \text{Bernoulli}(0.5), \quad j \in \{2, 4\}.
\end{aligned}$$

This data generating process follows the 1 and 2 assumptions.

[8]We also explore alternative scenarios in which the CATE distribution is bimodal, featuring two segments of equal or unequal proportions. The results are available in Web Appendix C.3.

We assume that the marketing intervention incurs a cost, c, which represents the decision threshold. In our simulation studies, we hold the (simulated) CATE distribution constant, with a mode and median equal to 1, and vary the intervention cost with $c \in \{0, 1, 2, 3\}$ to examine our approach's performance across different scenarios (see Figure 3 for an illustration).[9]

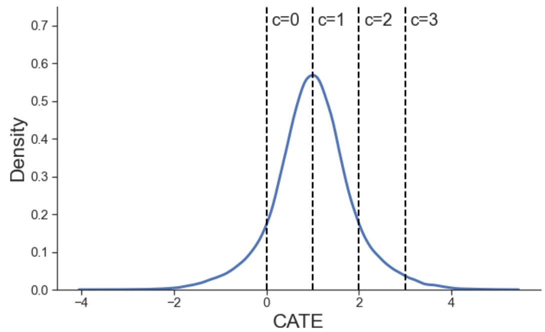


Figure 3: CATE Distribution and Intervention Costs
Each dashed line corresponds to a different intervention cost c. $c = 1$ corresponds to the mode of the CATE distribution. $c = 3$ represents the maximum deviation from the mode. $c = 0$ and $c = 2$ are symmetrically positioned around the mode.

Specifically, $c = 0$ represents scenarios with minimal intervention costs and positive impact; while such scenarios are rare in practice, an example might be an email campaign where the objective is simply for recipients to open the email. The case $c = 1$ represents scenarios where customers are predominantly clustered around the decision threshold. By contrast, $c = 3$ describes situations where only a few customers are near the decision threshold. This latter case is common in practice, either because treatment effects are generally low across customers or because intervention costs are prohibitively high (e.g., Ascarza et al. 2016; Lemmens and Gupta 2020).

Additionally, because $c = 0$ and $c = 2$ are symmetrically positioned around the mode of the CATE distribution (located at 1, making their distance to the mode equal), this setup allows us to investigate whether our approach's performance is influenced solely by the distance from the decision threshold to the mode, or also by the direction relative to the mode (with $c = 0$ below and $c = 2$ above). We also vary the experimental size (N_S) $\in$ {1k, 5k, 10k, 20k, 30k}) to assess sample size efficiency.

[9]Alternatively, we could vary the CATE distribution itself, but because only the difference between CATE and cost matters for the targeting decision, both approaches yield same insights.

5.2. Experimental Designs for Comparison

5.2.1. Policy-Aware Experimentation

We examine two variants of our proposed policy-aware approach: one employing multiple batches, with each batch comprising the same number of customers (in this case, 200 customers), and another utilizing only two batches of varying size.[10] In both cases, the total experimental size remains constant; the only difference is the number of batches drawn, which has implications for the number of times the researchers need to observe consumer outcomes and the number of times the EPL sampling strategy is employed. This comparison enables us to evaluate the performance differences between a more complex adaptive design and a simplified two-stage design, thereby elucidating the potential trade-offs between implementation complexity and targeting efficacy.[11]

5.2.2. Benchmarks

We compare our approach with three alternative experimental designs. The first is the test-then-learn approach (`Default`) commonly employed in practice (e.g., Ascarza 2018; Yang et al. 2023; Huang and Ascarza 2024). This method involves sampling customers with equal probability for the experiment and assigning them to different treatment arms at random. In our simulation, we assign the sampled customers to the two treatment arms with a probability of 0.5.

The second benchmark is the uncertainty sampling approach (`Uncertainty`) commonly used in active learning (e.g., Burbidge et al. 2007; Shankaranarayana 2023). This approach focuses on querying points with the highest estimation uncertainty:

$$S_b = \arg\max \sigma(\hat{\tau}_{S^{b-1}}(x))$$

where $\sigma(\hat{\tau}_{S^{b-1}}(x))$ denotes the standard deviation of the CATE estimate from previous batches. We choose uncertainty sampling over classical A- and D-optimal designs (e.g. Kiefer and Wolfowitz 1959; Fontaine et al. 2020) for two key reasons. First, A- and D-optimal designs target global parameters like average treatment effects, while our focus is on local heterogeneous effects (CATEs). Second, these designs are tailored to parametric models, whereas uncertainty sampling

[10]In the main text, we employ equal-sized batches (200 samples each) following Waisman et al. (2024) for the multi-stage design, with the number of batches varying by experimental size. In Appendix C.2.1, we evaluate alternative allocation schemes under a fixed number of batches (10) across different experimental sizes: both a constant batch size approach (where batch size varies with total experimental size) and a decreasing batch size configuration.

[11]For the two-stage design, we investigate various proportions of customers r to be sampled in the first stage ($r \in \{0.5, 0.7, 0.9\}$) in Appendix C.2.2, aiming to examine the optimal two-step configuration for different scenarios.

naturally accommodates local non-parametric estimators such as Causal Forest. This alignment with our methodological framework makes uncertainty sampling a more appropriate benchmark for evaluating our proposed approach.

The third benchmark is the state-of-the-art adaptive experimental design (`Adaptive`) proposed by Kato et al. (2024). While most adaptive designs in the literature focus on different objectives (e.g., balancing exploration and exploitation) or require strong parametric assumptions, Kato et al. (2024) aims to estimate the most accurate targeting policy with limited parametric assumptions. This alignment in objectives and assumptions makes it an ideal method for comparison with our approach. Note that, by design, this adaptive approach determines treatment assignment decisions for sampled customers rather than selecting which customers to include in the experiment. Specifically, the Kato et al. (2024) approach assigns the customers sampled in batch b to different treatment arms based on the following rule:

$$P_{\mathcal{S}^{b-1}}(W_i = 0|\mathbf{X}_i = x) = \frac{\sigma^0_{\mathcal{S}^{b-1}}(x)}{\sigma^0_{\mathcal{S}^{b-1}}(x) + \sigma^1_{\mathcal{S}^{b-1}}(x)}$$

$$P_{\mathcal{S}^{b-1}}(W_i = 1|\mathbf{X}_i = x) = \frac{\sigma^1_{\mathcal{S}^{b-1}}(x)}{\sigma^0_{\mathcal{S}^{b-1}}(x) + \sigma^1_{\mathcal{S}^{b-1}}(x)}$$

where $\sigma^w_{\mathcal{S}^{b-1}}(x)$ denotes the standard deviation of the potential outcomes $Y_i(W_i = w)$ estimated from the previous $b-1$ batches.[12] Intuitively, this approach prioritizes assigning customers to the treatment arm with greater uncertainty in expected outcomes, allowing the firm to reduce uncertainty in customer responses and identify the most effective treatment for each customer more accurately. This is in contrast to our approach, where we selectively sample the customers whose prediction errors in CATE estimation have the most significant impact on the firm's profitability, and randomly assign them to different treatment conditions.

5.3. Evaluation Procedure

Each sampling strategy yields a targeting policy $\hat{\pi}_S(\mathbf{X}_i)$, which we evaluate using a separate held-out dataset, simulating future deployment of the learned policy. For each method, we perform 100 replications: in each replication, we run the experiment, estimate the CATE model, derive the targeting policy, and assess its performance on the validation set. We then report the average results across replications.

[12] To account for the adaptive nature of the experimental data, we follow Kato et al. (2024) and reweight the outcome Y_i for customer i sampled in batch b by $P_{\mathcal{S}^{b-1}}(W_i|\mathbf{X}_i)$. See Web Appendix C.1 for implementation details.

Specifically, we generate a customer base of $N_{\mathcal{I}} = 100{,}000$ and a fixed evaluation set of $N_{eval} = 10{,}000$ customers, used across all replications. In each replication, we sample N_S customers from the base to conduct the experiment, yielding experimental data S used to estimate the CATE model. This model is then used to generate targeting decisions $\hat{\pi}(\cdot)$ for the evaluation set D_{eval}.

We measure policy performance using the proportional profit gap (PPG), defined as:

$$\text{PPG} = \frac{\sum_{i \in D_{eval}} (\tau(\mathbf{X}_i) - c) \cdot \pi^*(\mathbf{X}_i) - \sum_{i \in D_{eval}} (\tau(\mathbf{X}_i) - c) \cdot \hat{\pi}_S(\mathbf{X}_i)}{\sum_{i \in D_{eval}} (\tau(\mathbf{X}_i) - c) \cdot \pi^*(\mathbf{X}_i)}$$

where $\pi^*(\cdot)$ denotes the true optimal policy based on the true CATE $\tau(\cdot)$. The PPG captures the relative loss in profit from using the estimated policy $\hat{\pi}(\cdot)$ instead of the optimal one. Lower values indicate better performance.

5.4. Results

5.4.1. Profitability of Targeting Policies

Figure 4 shows the proportional profit gaps of the targeting policies learned by different experimental designs across various intervention costs ($c \in \{0, 1, 2, 3\}$) and experimental sizes ($N_S \in \{1k, 5k, 10k, 20k, 30k\}$). Note that a smaller gap (indicating that the estimated policy is closer to optimal) is better. Across all experimental sizes, EPL-based designs generally achieve lower profit gaps than the three benchmarks, particularly when the intervention cost c lies in the tails of the CATE distribution. These gains are most pronounced at smaller sample sizes (e.g., $N_S = 1k$ or $5k$), where EPL more effectively concentrates learning on the most consequential customers. At larger sample sizes, EPL performs comparably to the Uncertainty benchmark. This performance pattern holds for both the multi-stage and the two-stage versions of our approach, with the latter offering similar gains despite its simpler structure.

To evaluate performance independently of the proportional metric's denominator, we also report absolute profit gaps in Figure 5. This comparison helps isolate performance patterns, particularly the observed symmetry between $c = 0$ and $c = 2$.

The results reveal several insights. First, when the decision threshold deviates from the mode of the CATE distribution (i.e., $c \in \{0, 2, 3\}$), our approach generally outperforms the three benchmarks, particularly at lower sample sizes. The similar absolute profit gaps between $c = 0$ and $c = 2$ suggest that performance is primarily driven by the distance between the decision thresh-

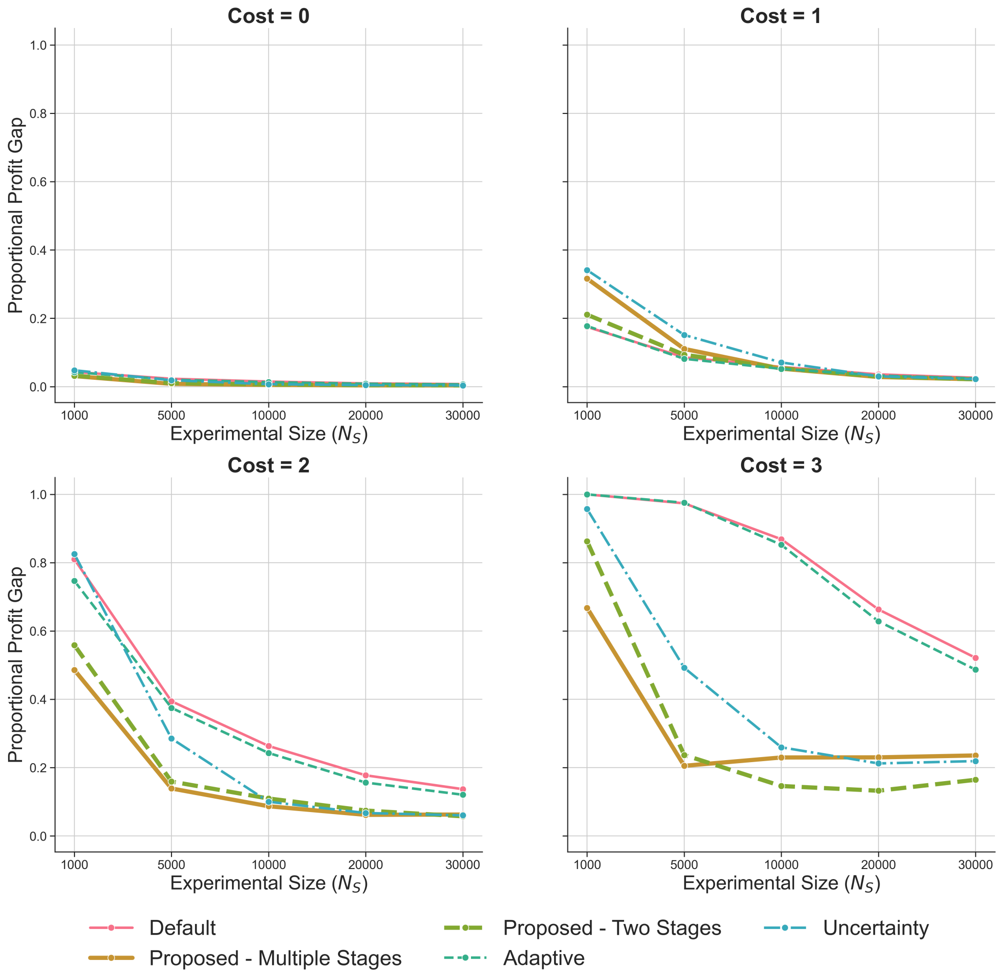


Figure 4: Proportional Profit Gaps of Different Experimental Designs
We report the average value of the proportional profit gap across 100 replications. Each line corresponds to an experimental approach. Lower values indicate better performance.

old and the mode of the CATE distribution. [13] In contrast, when the decision threshold aligns with the mode of the CATE distribution (i.e,. $c = 1$), our approach performs comparably to the `Default` and `Adaptive` approaches and may underperform at small sample sizes ($N_S < 10000$). In this case, random sampling naturally concentrates observations around the threshold, while our method's EPL estimates — still imprecise at low N_S — may lead to suboptimal prioritization. [14]

[13]The proportional profit gap is calculated relative to the profit generated by the optimal policy. Because $c = 0$ reflects a costless intervention, the optimal policy yields higher overall profit compared to $c = 2$, where the intervention is more costly. As a result, even when the absolute profit losses are similar, the proportional profit gap for $c = 0$ appears smaller due to the larger denominator. This distinction highlights why proportional and absolute metrics can tell different stories.

[14]Notably, our approach generally outperforms uncertainty sampling across all scenarios, especially at smaller sample sizes ($N_S \leq 10,000$). This supports the notion that reducing uncertainty alone is insufficient, sampling should also account for economic relevance.

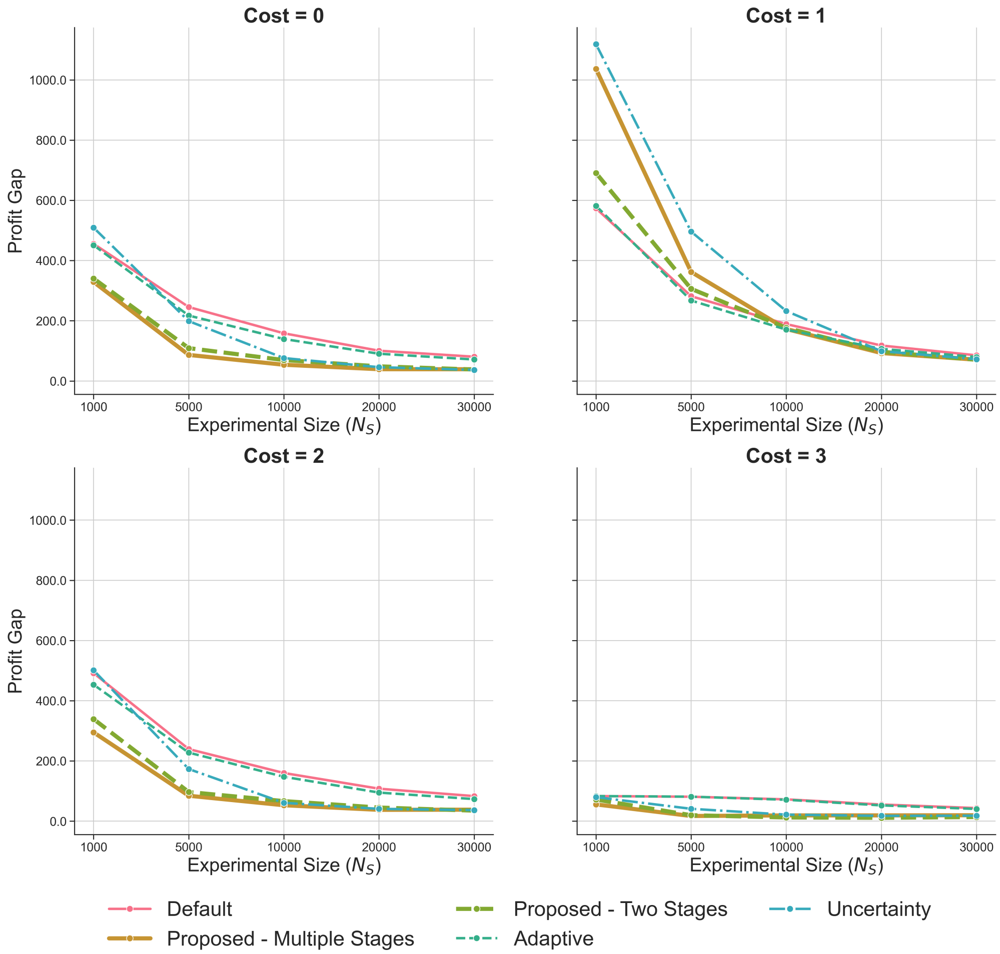


Figure 5: Profit Gaps of Different Experimental Designs
We report the average value of the profit gap (i.e., the difference between the incremental profit of $\hat{\pi}(X_i)$ and that of the optimal policy) across 100 replications. Each line corresponds to an experimental approach. Lower values indicate better performance.

Beyond offering empirical validation, these results clarify when policy-aware experimentation offers the most value. Specifically, when fewer customers lie near the targeting threshold, our approach enables firms to build more profitable policies. Such scenarios emerge when intervention costs are misaligned with central patterns in responsiveness—especially when: (1) the intervention is harmful for many customers (e.g., Ascarza et al. 2016), shifting treatment effects leftward against a positive cost; (2) the intervention is expensive (e.g., mailings, service calls) and effectiveness is low (e.g., Lemmens and Gupta 2020; Simester et al. 2022), pushing most treatment effects toward zero with a positive threshold; or (3) the intervention risks cannibalizing natural revenue (e.g., discounts or free goods) (e.g., Anderson and Simester 2004; Ailawadi et al. 2007; Ascarza

2018; Yoganarasimhan et al. 2023; Yang et al. 2023), which often generates negative CATEs. We examine this last scenario in detail in Section 6 using two field applications.[15]

Second, we emphasize the sample efficiency of our approach. In scenarios where the cost deviates from the mode of the CATE distribution (i.e., $c \in \{0, 2, 3\}$), our approach matches or exceeds the performance of benchmarks using far fewer observations. For example, when $c = 2$, our method requires only 5k samples to match the profit levels attained by other methods using 10k samples. This efficiency becomes even more pronounced as the threshold shifts further from the mode (i.e., $c = 3$), reinforcing the cost-saving potential of our design.

Finally, we underscore the effectiveness of our approach even in its simplified two-stage design. Across settings where $c \in \{0, 2, 3\}$, the two-stage version consistently outperforms all benchmarks and performs comparably to the fully adaptive version. This demonstrates that most of the benefits of policy-aware experimentation can be achieved without the operational complexity of full adaptivity, making the method highly practical for real-world deployment.

5.4.2. Effectiveness in Identifying Consequential Customers

We further analyze the CATE distributions of the (sampled) customers to understand the effectiveness of each of the approaches in identifying the most consequential customers. Unlike real-world data, the simulation setting allows us to observe the true CATEs and therefore offers the opportunity to directly assess whether our sampling approaches successfully identify and select those customers whose CATE estimation errors are most consequential for targeting profitability.

Figure 6 illustrates the CATE distributions of the customers sampled by different approaches (`Default`, `Uncertainty`, `Proposed`) together with the distribution of the entire population (`Population`).[16] Each subfigure presents one of the scenarios, depending on the intervention cost.[17]

Consistently across all scenarios, the EPL criterion (`Proposed`) selects customers whose CATEs are close to, yet slightly deviate from the decision threshold. Notably, when the cost aligns with the mode of the CATE distribution (i.e., $c = 1$), the CATE distribution of customers selected by our (`Proposed`) approach is even more concentrated around this mode than those selected by random sampling. Additionally, when the cost diverges from the mode (i.e., $c \in \{0, 2, 3\}$), our method re-

[15]Although our approach performs well even when $c = 0$ and CATEs are positive, such settings often render a simple "target everyone" rule nearly optimal, leaving little room for improvement through refined targeting.

[16]Since the `Adaptive` approach relies on random sampling of customers, its results are identical to those of the `Default` approach.

[17]Here we set the sample size to 10k and implement the EPL sampling strategy using a two-stage design with $r = 0.5$.

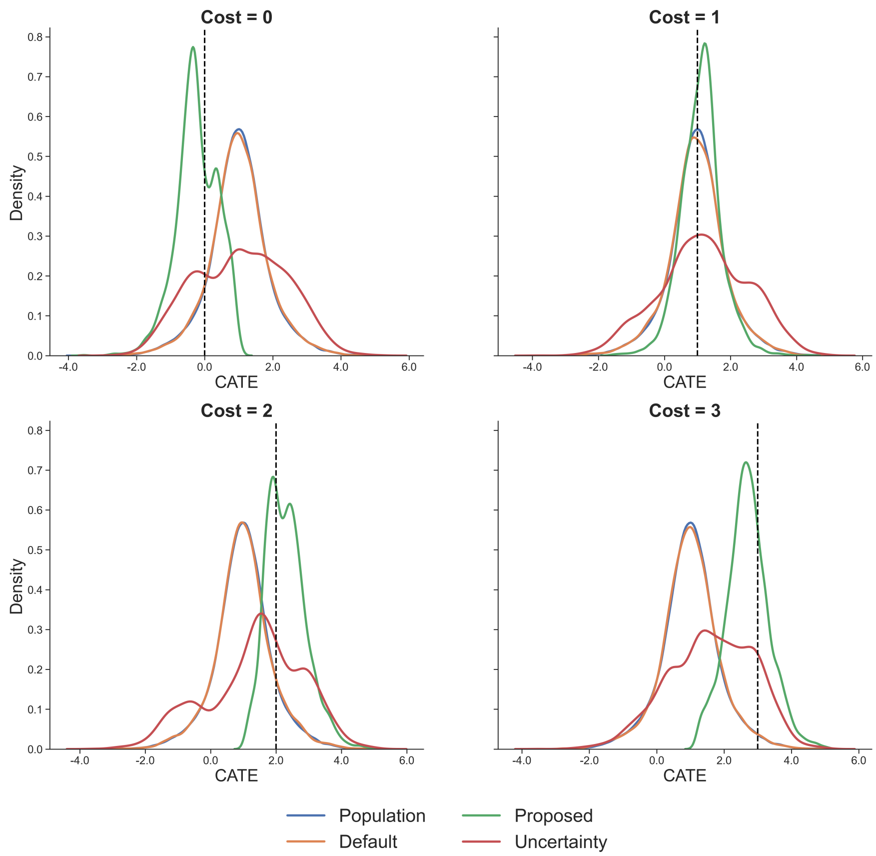


Figure 6: CATE Distributions of Customers Sampled by Different Approaches
Each line corresponds to the CATE distribution of the customers sampled by different approaches. The dashed line represents the intervention cost, which is also the decision threshold.

sults in a CATE distribution that concentrates around, though not directly at, the decision threshold.[18]

These results underscore the effectiveness of our approach in pinpointing consequential customers — those with CATEs near but not exactly at the decision threshold — and sampling them intensively. Conversely, customers selected through random sampling (`Default`) more closely resemble the population distribution, while `Uncertainty` sampling produces a more dispersed distribution, oversampling extreme values at the expense of the profitable boundary regions. Consequently, the distinction in the sampling of consequential customers between our approach and the benchmarks becomes particularly pronounced when a smaller fraction of customers are positioned near the decision threshold. This finding emphasize the advantage of our method in boost-

[18]When $c = 3$, our method tends to focus on sampling customers whose CATEs are just below the threshold, as the pool of customers with CATEs above the threshold is too limited for intensive sampling in this case. By contrast, when $c = 2$, there are a larger number of customers with CATEs above the threshold, enabling us to sample these customers more intensively as they are less likely to have been sampled in the first stage. Similarly, for $c = 0$, our approach intensifies sampling in regions where customers are less likely to have been selected initially.

ing targeting performance under conditions with limited customers around the decision threshold.

In conclusion, our simulation results affirm the superiority of our approach in enhancing targeting performance and sample size efficiency, especially in scenarios where a relatively small fraction of customers are positioned around the decision threshold. This advantage arises from our method's ability to effectively identify consequential customers and prioritize their sampling, an outcome unattainable with random sampling alone. Furthermore, we show that even a simplified two-stage design of our approach maintains its efficacy, providing firms with a practical, efficient solution for improving targeting outcomes in cases where a fully adaptive design may be impractical. Taken together, the simulation findings validate the theoretical foundations of our method and highlight its potential to improve both targeting precision and experimental efficiency, particularly when customers near the decision threshold are scarce. We next assess whether these gains persist in more complex and realistic environments by turning to two empirical applications.

6. Empirical Applications

6.1. Overview

We evaluate our proposed method using two real-world marketing campaigns: a dormant-customer reactivation effort by a telecommunications firm, and a mobile app-based promotion run by a national coffee chain. Across both settings, we compare the performance of our approach against baseline and benchmark strategies under varying experimental sizes and cost structures. In both cases, we evaluate the targeting performance of three types of targeting policies, the `Default` targeting policy, which is learned from randomly sampling customers for the experimental set and observing their outcomes, the `Uncertainty` targeting policy learned from the observed outcomes of the customers sampled by the uncertainty approach, and the `Proposed` policy, which is learned by observing the outcomes of the customers selected via our proposed sampling strategy.[19] Consistent with the analyses in previous sections, we evaluate different variants of our approach by

[19] Given that the data has already been collected, we do not include Kato et al. (2024) as a benchmark in the empirical application. The primary reason is that the optimal treatment assignment rule from Kato et al. (2024) might allocate customers to different treatment conditions than those in the original experimental data. Consequently, implementing Kato et al. (2024) offline would necessitate accurately simulating customers' counterfactual behavior. Given the low response rates in both datasets, accurately predicting customers' counterfactual behavior in these contexts is very challenging.

modifying the number of stages: from multiple batches (each with a size of 500) to a simplified two-stage design.[20] Finally, we test the performance of the different approaches across varying experimental sample sizes N_S.

Although our theoretical guarantees apply to a single-shot design with a sufficiently large initial sample, our empirical applications adopt a sequential implementation to reflect practical constraints. This setup mirrors the simulation framework and enables progressive refinement of EPL estimates as more data become available.

To assess the performance of each targeting approach, we use the bootstrap validation scheme similar to Ascarza (2018), where we generate 100 data splits, with each split consisting of a customer base, $\mathcal{I}$; 80%, and an evaluation set, D_{eval}; 20%. In each split, we sample N_S customers from the customer base, $\mathcal{I}$, following each of the sampling criteria (`Default`, `Uncertainty` and `Proposed`). These sampled customers would constitute the experimental data, S, and will, therefore, be the data used to estimate the CATE model (for each approach). We then leverage the constructed CATE model to generate targeting decisions $\hat{\pi}(\cdot)$ on the evaluation set, D_{eval}.

We evaluate the targeting performance by first computing the expected profit generated by each estimated targeting policy, $\hat{\pi}_S(\cdot)$, as well as a uniform policy, π_u, that provides the treatment to every customer. Specifically, we use the inverse-probability-weighted (IPW) estimator (Horvitz and Thompson 1952; Hitsch et al. 2024) to estimate the expected profit of a targeting policy, $\hat{\pi}(\cdot)$:

$$\text{Profit}(\hat{\pi}) = \sum_i \left(\frac{1 - W_i}{1 - e(\mathbf{X}_i)} (1 - \hat{\pi}_S(\mathbf{X}_i)) Y_i(0) + \frac{W_i}{e(\mathbf{X}_i)} \hat{\pi}_S(\mathbf{X}_i)(Y_i(1) - c_i) \right) \tag{8}$$

where $e(\mathbf{X}_i)$ is the (estimated) propensity score of customers assigned to the treatment condition in the evaluation set D_{eval}.

After obtaining the expected profit estimates, the *proportional profit improvement* of the targeting policy $\hat{\pi}(\cdot)$ relative to the uniform policy π_u is then computed as the evaluation metric using the following formula:

$$\text{PPI} = \frac{\text{Profit}(\hat{\pi}_S) - \text{Profit}(\pi_u)}{\text{Profit}(\pi_u)} \tag{9}$$

where $\text{Profit}(\pi_u)$ denotes the expected profit of the uniform policy. Since this metric quantifies the profit improvement of the estimated targeting policies relative to the uniform policy, a larger PPI indicates that the targeting decision estimated from the experimental data D_e is more profitable.

[20] For the multi-stage design, we follow a similar procedure as Waisman et al. (2024) by using equal-sized batches throughout the experiment. We investigate alternative sample allocation schemes for the multi-stage design and explore different configurations for the two-stage design in Appendix D.4.

6.2. Telecommunication Dormant Reactivation Campaign

6.2.1. Empirical Context and Data Description

Our first empirical application leverages the data from reactivation campaign of dormant customers conducted by a telecommunication company. The experimental data involves a reactivation campaign aiming to activate and increase users' usage over a 14-day period. The experiment ran from April 7^{th}, 2016 to January 10^{th}, 2018. Each week, the company identified eligible customers and randomly assigned them to either the control or treatment group. The campaign included 374,051 eligible customers, with 83,781 in the control group and 290,270 in the treatment group. Customers in the treatment group were offered 5 units of international voice credit for 3 days if they recharged at least 20 credit units within 3 days of receiving the offer. The outcome variable, a continuous measure, represented the total expenditure of each customer over the subsequent 14 days. While the intervention was cost-free for the company (i.e., giving free international voice credits doesn't cost the company anything), offering free international voice credits creates cannibalization as some of the customers would have otherwise paid for these credits. This cannibalization resulted in treated customers recharging, on average, less than those in the control group, with a negative average treatment effect of $-.65$, corresponding to a 5.8% decrease in average recharge in the next 14 days.

The dataset also includes various pre-treatment customer behaviors, which serve as covariates for targeting. These variables capture customers' previous behavior such as usage, recharge activity, and cancellation activity (of related services) over the past 7, 14, and 30 days. Additionally, because the focal company ran the campaign on a weekly basis, we created an additional variable, targeting ratio, which represents the proportion of customers targeted in a given week and is added as a control in all models. See Web Appendix D.1 for further details about the data.

Our dataset involves a series of experiments run every week, each with a random set of customers. Recall that the proposed approach should be particularly useful when there is greater misalignment between the CATE distribution and the decision threshold. In our data, this should occur when the intervention has a negative effect due to cannibalization of organic spending, leading to a high opportunity cost of the intervention. To assess this situation, we compare the analysis on the full dataset, with an analysis that focuses on the weeks with negative ATEs. The dataset of negative ATE weeks contains 40.7% of weeks, including 149,411 customers, with an average treatment effect of -2.02, demonstrating substantially greater misalignment between the CATE distribution and the decision threshold compared to the full dataset. By comparing performance across the full dataset and the partitioned data, we can assess how our approach performs under different relationships between the CATE distribution and the decision threshold, similar to the analysis presented in Section 5.

6.2.2. Profitability of Targeting Policies

We now turn to evaluate the performance of the different policies. Recall that for this analysis, we split the data into customer base, $\mathcal{I}$, used to run each experimental approach (default or proposed), and evaluation set, D_{eval}, used to evaluate the profitability of each policy, as defined in Equation (9). Figure 7 shows the proportional profit improvement of the targeting policies relative to the uniform policy. Each line corresponds to the profitability of the policy derived by each experimental design (or sampling approach), across experimental sizes ($N_S \in \{10k, 20k, 30k, 40k, 50k\}$). In addition, we report the percentage of replications in which the proportional profit improvement of the focal method relative to the uniform policy is greater than the default approach in Table 1 to provide a better sense of the performance difference across the different approaches.

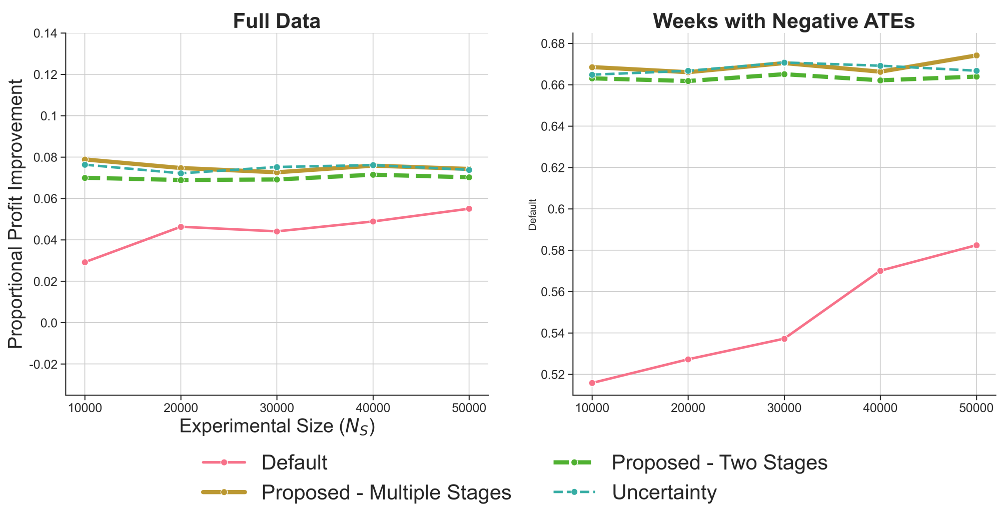


Figure 7: Performance of Targeting Policies Learned from Different Experimental Designs (Telecommunication)

We report the average value of the proportional profit improvement relative to the uniform policy across 100 replications. Each line corresponds to an experimental approach. Higher values indicate better performance.

The results reveal that our approach significantly outperforms the `Default` approach and performs comparably to the `Uncertainty` approach.[21] The performance advantage over the default

[21] The comparable performance with uncertainty sampling reflects substantial overlap in customer selection; approximately 80% of sampled customers are identical across both methods.

Table 1: Percentage of Replications in which the Proposed Method Outperforms the Default Approach (Telecommunication)

Experimental Design	Experimental Size				
	10k	20k	30k	40k	50k
Panel A: Full Data					
Proposed Multiple Stages	74%	61%	69%	66%	59%
Proposed Two-Stage	69%	62%	68%	63%	50%
Uncertainty Sampling	78%	62%	65%	67%	64%
Panel B: Weeks with evidence of cannibalization (Negative ATEs)					
Proposed Multiple Stages	72%	79%	80%	75%	80%
Proposed Two-Stage	73%	77%	79%	76%	75%
Uncertainty Sampling	69%	78%	80%	76%	81%

Note: We report the percentage of bootstrap replications in which the proportional profit improvement of the focal method relative to the uniform policy is greater than the default approach.

approach is even more pronounced in the partitioned data with negative ATEs, consistent with our simulation findings that greater misalignment between CATE distribution and decision threshold amplifies our method's relative improvement. Remarkably, our method requires only 10,000 experimental samples to generate a more profitable targeting policy than the default method achieves with 50,000 samples. This demonstrates the practical value of our approach, especially in scenarios where increasing the experimental size is difficult or costly for firms, such as when the customer base is limited.

Additionally, the simplified two-stage designs achieve similar profitability to that of the fully adaptive design, regardless of the experimental size. Our findings highlight that even with a simplified implementation, the proposed method can enhance profitability, making it a valuable tool for firms with different capacities for experimental scale.

6.3. Starbucks Promotional Campaign

6.3.1. Empirical Context and Data Description

Our second empirical application utilizes the data from a promotional campaign conducted through Starbucks' mobile reward app.[22] The experimental data involves a promotional campaign aiming to increase customers' purchase rate. In particular, the dataset contains 126,184 customers who were randomly assigned to either the control group (63,112) or the treatment group (63,072), with the treated customers receiving the promotional content offered by Starbucks. The outcome variable is binary, indicating whether the customer made a purchase or not. Notably, the response rates to the intervention are quite low, with 1.68% in the treatment group and .73% in the control group, leading to an average treatment effect of .95%. The data also includes seven pre-treatment covariates, which will be used for targeting.[23] See Web Appendix D.2 for further details about the data.

Because the data provided does not include specific details about the promotional content, we consider two scenarios. First, one in which Starbucks sent a promotional content without discount (e.g., a notification promoting Starbucks). This is the same as analyzing an intervention with, essentially, no cost or potential cannibalization. Second, we assume that the intervention sent to the treatment group was offering a 50% off discount in the next purchase. In this scenario, the intervention creates cannibalization since using the promotion implies lower revenue in the associated sale.[24] In both scenarios, we assume that each purchasing customer has an order value of $6, which is the average ticket size for a Starbucks customer.[25]. However, for customers who receive the 50% discount, their spending is reduced to $3 per order. Importantly, each scenario reflects different levels of cannibalization introduced by the intervention, enabling us to assess our approach's performance under varying relationships between the CATE distribution and the decision threshold, similar to our analysis in Section 5.

[22]The data is provided by Starbucks and was made available through the Udacity Data Science Program.

[23]The provided data is anonymized and does not include the meaning of each pre-treatment variable. While this limits the ability to interpret some findings, these variables contain the necessary information for determining targeting policies, which is our primary objective.

[24]We consider this case a cost-free intervention with a risk of cannibalization, rather than a costly intervention, due to the intervention's effect of reducing customer spending.

[25]Source: https://wifitalents.com/statistic/starbucks-customers/

6.3.2. Profitability of Targeting Policies

We now turn to evaluate the performance of the different policies. Similar to the first empirical application, we split the data into customer base, $\mathcal{I}$, used to run each experimental approach (`Default` or `Proposed`), and evaluation set, D_{eval}, used to evaluate the profitability of each policy, as defined in Equation (9). Figure 8 shows the proportional profit improvement of the targeting policies relative to the uniform policy for each scenario (no discount and 50%-off discount). Each line corresponds to the profitability of the policy derived by each experimental design (or sampling approach), across experimental sizes ($N_S \in \{30\text{k}, 35\text{k}, 40\text{k}, 45\text{k}, 50\text{k}\}$).

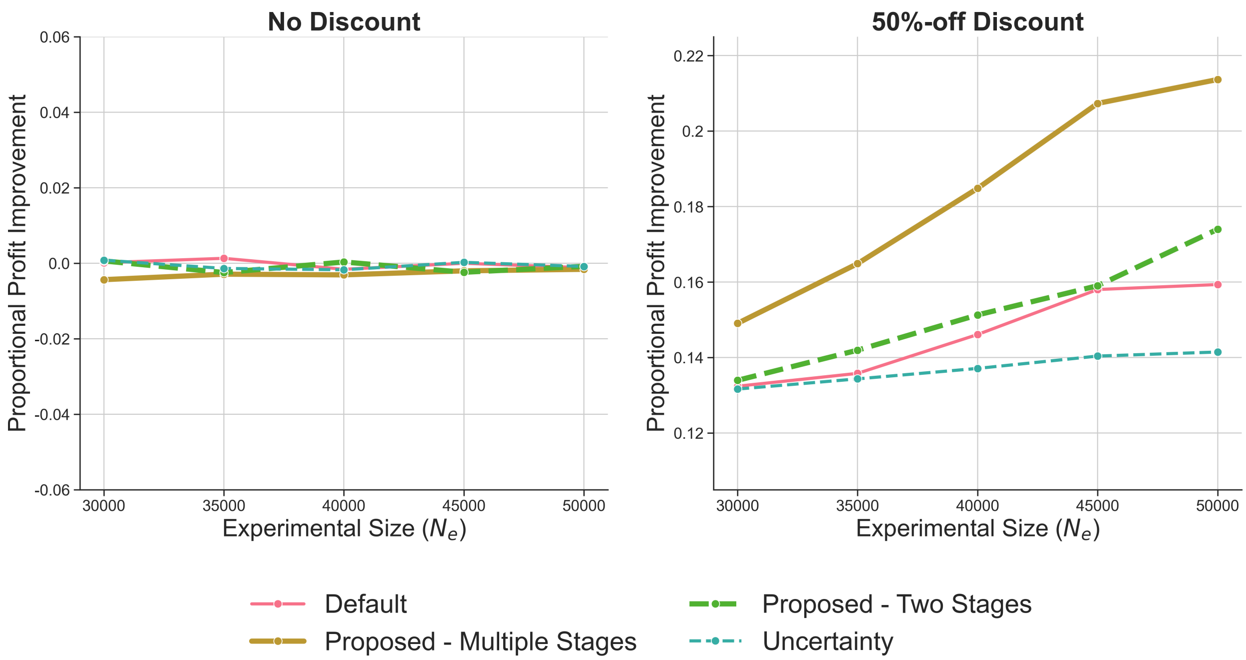


Figure 8: Performance of Targeting Policies Learned from Different Experimental Designs (Starbucks)
We report the average value of the proportional profit improvement relative to the uniform policy across 100 replications. Each line corresponds to an experimental approach. Higher values indicate better performance.

First, our approach significantly outperforms both benchmarks when Starbucks offers a 50%-off discount (i.e., when the number of customers around the decision threshold is limited) while it performs at par with the benchmark approach in the "No Discount" case. This finding is consistent with our simulation results and underscores the practical benefit of our method in enhancing the profitability of targeting policies, especially in scenarios where the intervention cost is high, but customer responsiveness might be low, as is the case when Starbucks offers a 50%-off discount.[26] Table 2 reports, for the scenario where Starbucks offer a 50%-off discount, the percentage of replications in which the proportional profit improvement is greater than the default approach.

Table 2: Percentage of Replications in which the Proposed Method Outperforms the Default Approach (Starbucks 50%-off Discount)

Experimental Design	**Experimental Size**				
	30k	35k	40k	45k	50k
Proposed Multiple Stages	57%	75%	76%	89%	93%
Proposed Two-Stage	49%	54%	53%	54%	55%
Uncertainty Sampling	48%	51%	40%	30%	29%

Note: We report the percentage of bootstrap replications in which the proportional profit improvement of the focal method relative to the uniform policy is greater than the default approach when Starbucks offers a 50%*-off discount.*

Second, even when the proposed approach does not provide additional benefits in learning targeting policies, as in Scenario 1, our method (whether fully adaptive or performed in two stages) does not harm profitability. This highlights the low risk of implementing the proposed approach in practice.

Third, when comparing the performance of targeting policies across different experimental sizes, the default method requires at least $50k$ experimental samples to achieve the same level of profitability as the fully adaptive method with only $35k$ experimental samples. This underscores the value of the proposed approach, particularly in situations where increasing the experimental size is challenging or costly for firms, such as when the customer base size is limited.

6.4. Discussion of Results

Across both empirical applications, our proposed method consistently shows either comparable or superior performance relative to both benchmarks. Notably, it excels when the intervention creates a risk of cannibalizing profits that would have been earned in the absence of intervention, as evidenced in both the telecommunications application (offering free credits) and the 50% discount scenario for Starbucks (offering discounts). In these cases, the interventions reduce incremen-

[26] In the "No Discount" scenario, the estimated targeting policies perform similarly to the uniform policy across different experimental designs. This suggests that when the intervention is cost-free and customers respond positively, a simple strategy of targeting all customers would already be close to an optimal approach.

tal spending for targeted customers, resulting in a negative average treatment effect. This result arises because, with a combination of weak treatment effects and cannibalization caused by the intervention, the CATE distribution tends to be leftward shifted (negative treatment effect). However, the firm's decision threshold still falls at zero. This misalignment between the mode of the CATE distribution and the decision threshold results in a smaller fraction of customers positioned near the threshold, making it harder to identify critical customers effectively.

Notably, this scenario frequently occurs in various marketing campaigns involving free goods or monetary incentives like coupons (e.g., Ascarza 2018; Yang et al. 2023). In addition, as predicted by our simulation results, our approach is also expected to be beneficial when there is a misalignment between intervention costs and the central tendency of customer responsiveness, such as when (1) the intervention is detrimental for most customers (e.g., Ascarza et al. 2016), or (2) costly interventions (e.g., phone calls, mailings) show low effectiveness (e.g., Lemmens and Gupta 2020). In general, we recommend firms adopt our approach whenever there is a likely disparity between intervention cost and customer responsiveness.[27]

7. Conclusion

This paper develops a novel experimental design framework, *policy-aware experimentation*, that aligns experimental sampling directly with firms' profit-maximizing targeting goals. Instead of selecting experimental samples uniformly at random, our approach adaptively prioritizes customers whose treatment effect estimation errors are most likely to impact profitability. We introduce a new criterion, *expected profit loss (EPL)*, and show that sampling based on this criterion yields a near-optimal approximation to the globally optimal sampling strategy that yields the profit-maximizing targeting policy.

To operationalize this idea, we propose a sequential design guided by the EPL sampling strategy and introduce an estimation strategy that leverages the power of Bayesian inference in uncertainty quantification via Causal Forests. This enables firms to identify and oversample "consequential" customers — those near the decision threshold with high uncertainty — without requiring knowledge of true treatment effects. The method can be implemented in either a multi-stage

[27] Theoretically, our approach is also expected to perform well when the intervention is costless and customers are highly responsive, as the decision threshold in such cases similarly deviates from the mode of the CATE distribution. However, in these scenarios, a straightforward strategy that targets every customer would already be close to optimal, offering limited room for further improvement in targeting performance as shown in Section 6.3.2.

or a simplified two-stage format, depending on practical constraints such as delayed feedback or operational costs.

Our simulations and empirical analyses demonstrate that policy-aware experimentation offers substantial improvements over conventional sampling strategies — including standard test-then-learn, uncertainty sampling, and a leading adaptive strategy — especially in contexts where only a limited share of customers lies near the decision threshold. These settings frequently occur when interventions are broadly harmful (e.g. Ascarza et al. 2016), expensive with uncertain returns (e.g. Lemmens and Gupta 2020), or pose a risk of revenue cannibalization (e.g. Anderson and Simester 2004) — conditions under which traditional sampling fails to prioritize high-impact customers.

Managerial Implications. For marketing leaders aiming to improve the ROI of experimental targeting, our findings offer actionable guidance. The expected profit loss (EPL) sampling strategy is particularly valuable when a misalignment exists between intervention costs and the central tendency of customer responsiveness — conditions under which uniform sampling often fails to prioritize the right customers. EPL overcomes this by directing experimental resources toward individuals whose targeting decisions are most economically consequential. In our empirical applications, the proposed method improves profit outcomes by 5% to 10% compared to standard sampling approaches, and delivers similar results with up to 80% fewer experimental observations — underscoring its efficiency in data usage and impact on decision quality. This makes the approach especially attractive in constrained environments where expanding experimental size is difficult or costly. Moreover, EPL sampling can be implemented using off-the-shelf CATE estimators such as Causal Forests in EconML, reducing engineering complexity and enabling rapid deployment. Even a simplified two-stage implementation retains these benefits, minimizing operational burden and accelerating learning in settings with delayed feedback, such as promotions, churn prevention, or cross-channel campaigns.

Limitations and Future Directions. Although our research provides a simple and efficient solution for firms to improve targeted policies, there are limitations that suggest promising directions for future research. First, we focus on scenarios where the firms' profit-maximizing objectives are not subject to any constraints. However, in practice, some firms may face several managerial constraints on their targeted policies, such as budget constraints or fairness constraints (Lu et al. 2023). Future research could build on the decision-aware learning literature (e.g. Chung et al. 2022; Liu et al. 2023) to explore different sampling strategies that integrate firms' business objectives into the experimental design while accounting for these constraints.

Second, our approach assumes that firms lack prior knowledge about consumers' responsiveness to interventions. However, firms often possess historical experimental or observational data that, while related, may differ from the current experiment. For instance, firms may have conducted campaigns involving different interventions, such as varying discount levels. While previous research has explored transferring knowledge from past marketing campaigns to enhance targeting policies (Timoshenko et al. 2020; Huang et al. 2024) it has not addressed the misalignment between experimentation approaches and the firm's objectives. Future research could investigate ways to incorporate information from previous campaigns to refine the design of focal experiments in a policy-aware manner.

Third, although our two-stage design remains effective when there is a delay between the intervention and the outcome of interest, more efficient methods could address delayed feedback. Prior work has examined strategies for handling delayed feedback in multi-armed bandits by explicitly modeling the relationship between partial and delayed feedback (e.g. Grover et al. 2018). Building on these approaches, future research could extend our proposed sequential design to better manage significantly delayed feedback and reduce the overall lag time.

Fourth, our approach assumes that firms have a fixed customer base, allowing them to easily determine which customers warrant closer attention based on the potential impact of CATE estimation errors on profitability. While this scenario is common in various marketing applications such as customer retention (e.g., Ascarza 2018; Lemmens and Gupta 2020; Yang et al. 2023) and most promotional activities to incentivize consumption (e.g., Hitsch et al. 2024; Simester et al. 2020), there are instances where customers arrive sequentially and unpredictably, such as in digital advertising or customer acquisition strategies. In these cases, strategic sampling becomes more challenging, as firms cannot anticipate whether future customers might be more consequential and thus deserve greater focus. Future research could draw upon insights from the online active learning literature (Cacciarelli and Kulahci 2024) to investigate optimal sampling strategies in such dynamic environments.

Finally, alternative enhancements to our policy-aware approach could be explored. For instance, we proposed selectively sampling customers while assigning treatments randomly. Future research could investigate methods for both selectively sampling customers and strategically assigning them to treatment conditions to better align with firms' business objectives and improve targeting performance. Additionally, in cases where control condition is equivalent to business as usual, one could extend our work to use all non-experimental customers as a control group to enhance the estimation efficiency. Future research could explore the optimal sampling strategy under this framework.

Overall, our research demonstrates the value of incorporating firms' business objectives into the design of experiments. We hope that our work will inspire further research on aligning the science of experimentation with firms' objectives across a wider range of empirical and operational contexts.

References

Ailawadi, K. L., Harlam, B. A., César, J., and Trounce, D. (2007). Practice Prize Report—Quantifying and Improving Promotion Effectiveness at CVS. *Marketing Science*, 26(4):566–575.

Anderson, E. T. and Simester, D. I. (2004). Long-Run Effects of Promotion Depth on New Versus Established Customers: Three Field Studies. *Marketing Science*, 23(1):4–20.

Ascarza, E. (2018). Retention Futility: Targeting High-Risk Customers Might be Ineffective. *Journal of Marketing Research*, 55(1):80–98.

Ascarza, E., Iyengar, R., and Schleicher, M. (2016). The Perils of Proactive Churn Prevention Using Plan Recommendations: Evidence from a Field Experiment. *Journal of Marketing Research*, 53(1):46–60.

Athey, S., Chetty, R., Imbens, G. W., and Kang, H. (2019a). The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely. Working Paper 26463, National Bureau of Economic Research.

Athey, S., Tibshirani, J., and Wager, S. (2019b). Generalized random forests. *The Annals of Statistics*, 47(2).

Athey, S. and Wager, S. (2021). Policy Learning With Observational Data. *Econometrica*, 89(1):133–161.

Bubeck, S., Munos, R., and Stoltz, G. (2010). Pure Exploration for Multi-Armed Bandit Problems. *arXiv preprint arXiv:0802.2655*. arXiv:0802.2655 [cs, math, stat].

Burbidge, R., Rowland, J. J., and King, R. D. (2007). Active Learning for Regression Based on Query by Committee. In Yin, H., Tino, P., Corchado, E., Byrne, W., and Yao, X., editors, *Intelligent Data Engineering and Automated Learning - IDEAL 2007*, volume 4881, pages 209–218. Springer Berlin Heidelberg, Berlin, Heidelberg. Series Title: Lecture Notes in Computer Science.

Cacciarelli, D. and Kulahci, M. (2024). Active learning for data streams: a survey. *Machine Learning*, 113(1):185–239.

Chung, T.-H., Rostami, V., Bastani, H., and Bastani, O. (2022). Decision-Aware Learning for Optimizing Health Supply Chains. *arXiv preprint arXiv:2211.08507*. arXiv:2211.08507 [cs].

Ellickson, P. B., Kar, W., and Reeder, J. C. (2022). Estimating Marketing Component Effects: Double Machine Learning from Targeted Digital Promotions. *Marketing Science*, page mksc.2022.1401.

Feit, E. M. and Berman, R. (2019). Test & Roll: Profit-Maximizing A/B Tests. *Marketing Science*, 38(6):1038–1058.

Fernández-Loría, C. and Provost, F. (2022). Causal decision making and causal effect estimation are not the same… and why it matters. *INFORMS Journal on Data Science*, 1(1):4–16.

Fontaine, X., Perrault, P., Valko, M., and Perchet, V. (2020). Online A-Optimal Design and Active Linear Regression. arXiv:1906.08509 [stat].

Fu, Y., Zhu, X., and Li, B. (2013). A survey on instance selection for active learning. *Knowledge and Information Systems*, 35(2):249–283.

Grover, A., Markov, T., Attia, P., Jin, N., Perkins, N., Cheong, B., Chen, M., Yang, Z., Harris, S., Chueh, W., and Ermon, S. (2018). Best arm identification in multi-armed bandits with delayed feedback. *arXiv preprint arXiv:1803.10937*. arXiv:1803.10937 [cs, stat].

Hadad, V., Rosenzweig, L. R., Athey, S., and Karlan, D. (2021). Designing Adaptive Experiments. Practitioner's Guide, Stanford Business School.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York, New York, NY.

Hitsch, G. J., Misra, S., and Zhang, W. W. (2024). Heterogeneous treatment effects and optimal targeting policy evaluation. *Quantitative Marketing and Economics*, 22(2):115–168.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960.

Horvitz, D. G. and Thompson, D. J. (1952). A Generalization of Sampling Without Replacement from a Finite Universe. *Journal of the American Statistical Association*, 47(260):663–685.

Hu, Y., Zhu, H., Brunskill, E., and Wager, S. (2024). Minimax-Regret Sample Selection in Randomized Experiments. *arXiv preprint arXiv:2403.01386*. arXiv:2403.01386 [econ, stat].

Huang, T.-W. and Ascarza, E. (2024). Doing More with Less: Overcoming Ineffective Long-Term Targeting Using Short-Term Signals. *Marketing Science*, page mksc.2022.0379.

Huang, T.-W., Ascarza, E., and Israeli, A. (2024). Incrementality representation learning: Synergizing past experiments for intervention personalization. *Available at SSRN 4859809*.

Jain, L., Li, Z., Loghmani, E., Mason, B., and Yoganarasimhan, H. (2024). Effective Adaptive Exploration of Prices and Promotions in Choice-Based Demand Models. *Marketing Science*, page mksc.2023.0322.

Jesson, A., Tigas, P., van Amersfoort, J., Kirsch, A., Shalit, U., and Gal, Y. (2022). Causal-BALD: Deep Bayesian Active Learning of Outcomes to Infer Treatment-Effects from Observational Data. *arXiv preprint arXiv:2111.02275*. arXiv:2111.02275 [cs, stat].

Kallus, N. and Mao, X. (2023). Stochastic Optimization Forests. *Management Science*, 69(4):1975–1994.

Kasy, M. and Sautmann, A. (2021). Adaptive Treatment Assignment in Experiments for Policy Choice. *Econometrica*, 89(1):113–132.

Kato, M., Okumura, K., Ishihara, T., and Kitagawa, T. (2024). Adaptive Experimental Design for Policy Learning. *arXiv preprint arXiv:2401.03756*. arXiv:2401.03756 [cs, econ, stat].

Kiefer, J. and Wolfowitz, J. (1959). Optimum Designs in Regression Problems. *The Annals of Mathematical Statistics*, 30(2):271–294.

Lemmens, A. and Gupta, S. (2020). Managing Churn to Maximize Profits. *Marketing Science*, 39(5):956–973.

Liu, M., Grigas, P., Liu, H., and Shen, Z.-J. M. (2023). Active Learning in the Predict-then-Optimize Framework: A Margin-Based Approach. *arXiv preprint arXiv:2305.06584*. arXiv:2305.06584 [cs, math, stat].

Lu, H., Simester, D., and Zhu, Y. (2023). Optimizing Scalable Targeted Marketing Policies with Constraints. *Available at SSRN 4668582*.

Misra, K., Schwartz, E. M., and Abernethy, J. (2019). Dynamic Online Pricing with Incomplete Information Using Multiarmed Bandit Experiments. *Marketing Science*, 38(2):226–252.

Puha, Z., Kaptein, M., and Lemmens, A. (2020). Batch Mode Active Learning for Individual Treatment Effect Estimation. In *2020 International Conference on Data Mining Workshops (ICDMW)*, pages 859–866, Sorrento, Italy. IEEE.

Schwartz, E. M., Bradlow, E. T., and Fader, P. S. (2017). Customer Acquisition via Display Advertising Using Multi-Armed Bandit Experiments. *Marketing Science*, 36(4):500–522.

Shankaranarayana, S. M. (2023). Model Uncertainty based Active Learning on Tabular Data using Boosted Trees. Version Number: 1.

Simester, D., Timoshenko, A., and Zoumpoulis, S. (2022). A Sample Size Calculation for Training and Certifying Targeting Policies. *Available at SSRN 4228297*.

Simester, D., Timoshenko, A., and Zoumpoulis, S. I. (2020). Targeting Prospective Customers: Robustness of Machine-Learning Methods to Typical Data Challenges. *Management Science*, 66(6):2495–2522.

Sundin, I., Schulam, P., Siivola, E., Vehtari, A., Saria, S., and Kaski, S. (2019). Active Learning for Decision-Making from Imbalanced Observational Data. *arXiv preprint arXiv:1904.05268*. arXiv:1904.05268 [cs, stat].

Timoshenko, A., Ibragimov, M., Simester, D., Parker, J., and Schoar, A. (2020). Transferring information between marketing campaigns to improve targeting policies. Technical report, Working Paper.

Wager, S. and Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, 113(523):1228–1242.

Waisman, C., Nair, H. S., and Carrion, C. (2024). Online Causal Inference for Advertising in Real-Time Bidding Auctions. *Marketing Science*, page mksc.2022.0406.

Wilder, B., Dilkina, B., and Tambe, M. (2019). Melding the Data-Decisions Pipeline: Decision-Focused Learning for Combinatorial Optimization. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):1658–1665.

Yang, J., Eckles, D., Dhillon, P., and Aral, S. (2023). Targeting for Long-Term Outcomes. *Management Science*, page mnsc.2023.4881.

Yoganarasimhan, H., Barzegary, E., and Pani, A. (2023). Design and Evaluation of Optimal Free Trials. *Management Science*, 69(6):3220–3240.

Zhao, Y., Zeng, D., Rush, A. J., and Kosorok, M. R. (2012). Estimating Individualized Treatment Rules Using Outcome Weighted Learning. *Journal of the American Statistical Association*, 107(499):1106–1118.

Web Appendix

Policy-Aware Experimentation:
Strategic Sampling for Optimized Targeting Policies

These materials have been supplied by the authors to aid in the understanding of their paper.

A. Comparison of Methodologies in Literature

Table W-1: Comparison of Methodologies in Literature

Literature	Objective	Sampling Strategy	Sampling Space	Pre-treatment Covariates	Pre-Segmentation	Sequential	Examples
Feit and Berman (2019)	Calculate sample size	N/A	N/A	N	N	N	N/A
Simester et al. (2022)	Calculate sample size	N/A	N/A	Y	Y	N	N/A
Hu et al. (2024)	Learning minimax-regret policy	Stratified Sampling	Context	Y	Y	N	N/A
Optimal Experimental Design	Minimizing estimation uncertainty of parametric models	A- and D-optimal	Context	Y	N	N	Kiefer and Wolfowitz (1959), Fontaine et al. (2020)
Multi-Armed Bandit	Balancing exploration and exploitation	TS, UCB, etc.	Action	N	N	Y	Schwartz et al. (2017), Misra et al. (2019), ?
Contextual Bandit	Balancing exploration and exploitation	TS, UCB, etc.	Action	Y	N	Y	Hauser et al. (2009), Li et al. (2010), Caria et al. (2020), Aramayo et al. (2022), Athey et al. (2022)
Best Arm Identification (BAI)	Identify best treatment with minimum sample	TS, UCB, etc.	Action	N	N	Y	Bubeck et al. (2010), Chick and Frazier (2012), Grover et al. (2018), Kasy and Sautmann (2021)
Contextual BAI	Learning personalized policy with minimum sample	TS, UCB, etc.	Action	Y	N	Y	Jedra and Proutiere (2020), Carranza et al. (2023), Kato et al. (2024)
Waisman et al. (2024)	Reducing experimentation Cost for ad effect estimation	TS	Action	Y	N	Y	N/A
Active Learning for Supervised Learning	Optimize data collection for supervised learning	Uncertainty Sampling, BALD, etc.	Context	Y	N	Y	Fu et al. (2013), Wang and Ye (2015), Cardoso et al. (2017)
Active Learning for CATE Estimation	Optimize data collection for CATE estimation	EMCM, BALD	Context	Y	N	Y	Puha et al. (2020), Jesson et al. (2022)
Active Learning for Decision Making	Optimize observational data collection for decision making	Uncertainty Sampling	Context	Y	N	Y	Sundin et al. (2019)
Active Learning for Predict-then-Optimize	Optimize data collection for predict-then-optimize	Margin-based Sampling	Context	Y	N	Y	Liu et al. (2023)
Our approach	Learning profit-maximizing targeting policy with pre-determined sample size	Expected Profit Loss Sampling	Context	Y	N	Y	N/A

B. Proof of Proposition 1

Throughout the analysis, we consider a class of local nonparametric CATE estimators widely used in the marketing literature including nearest neighbor with growing k, kernel estimators, and honest forest estimators with sub-sampling such as Causal Forest (Wager and Athey 2018; Athey et al. 2019). We further impose the following regularity assumptions on the CATE estimators:

Assumption W-1 (Regularity Conditions)**.**

1. *(Lipschitz continuity of the CATE and cost function) The CATE function $\tau(\cdot) : \mathbf{X} \to \mathbb{R}$ and the cost function $\varsigma(\cdot) : \mathbf{X} \to \mathbb{R}$ are L-Lipshitz continuous, i.e., for all $x, x' \in \mathbb{R}^p$, there exists a finite constant L such that:*
$$|\tau(x) - \tau(x')| \leq L\|x - x'\|$$
$$|c(x) - c(x')| \leq L\|x - x'\|.$$

2. *(Error-density log-concavity) Let $Z = \frac{\hat{\tau}(x) - \tau(x)}{\sigma(x)}$ denote the standardized estimation error. For all $x \in \mathbf{X}$, the probability density function $f_Z(\cdot)$ is symmetric about zero and log-concave, i.e., $\log f_Z(u)$ is a concave function of $u \in \mathbb{R}$.*

3. *(Bounded influence region) Let d denote the dimensionality of the covariate space. There exists an $\alpha \in (0, \frac{1}{d})$ such that for all $n > |\mathcal{D}|$, the estimator at any query point $\mathbf{X}_j$ only uses observations within a ball of radius*
$$h_n = O(n^{-\alpha}).$$

4. *(Variance-shrinking property) For any $x \in \mathbf{X}$ and training set $S \in \mathcal{I}$, the pointwise variance of $\hat{\tau}(x)$ satisfies*
$$\sigma_S^2(x) = \frac{\sigma^2(x)}{\gamma_{|S|}(x)},$$
where $\sigma(x)$ is continuously differentiable and the sequence $\{\gamma_n(x)\}_{n=|\mathcal{D}|}^{\infty}$ with $\gamma_n(x) = nh_n^d$ satisfies the following properties:

   (a) *$\gamma_n(x)$ is monotonically non-decreasing for all n.*

   (b) *The sequence of first-order differences $\gamma_{n+1}(x) - \gamma_n(x)$ is monotonically non-increasing for all n.*

   (c) *For each n, the function $\gamma_n(\cdot) : \mathbf{X} \to \mathbb{R}^+$ is continuously differentiable.*

Note that the bandwidth and variance-shrinking property are generally satisfied for the class of CATE estimators we consider. The log-concavity condition encompasses a wide class of probability distributions frequently employed in the literature to model estimation errors, including the normal (Gaussian), logistic, Laplace, and Student's t-distributions with sufficient degrees of freedom.

We now prove Proposition 1:

Proof. To prove the approximation bound, we define the objective function as:

$$f(S) = \mathbb{E}\left[Y_i(0) + (\tau(\mathbf{X}_i) - c(\mathbf{X}_i)) \cdot \hat{\pi}_S(\mathbf{X}_i)\right],$$

where the expectation is taken w.r.t both the covariate distribution and the sampling distribution of $\hat{\pi}_S$. For notation simplicity, throughout the analysis, we denote $\mathbf{E}_{\mathbf{X}}$ as the expectation taken w.r.t the covariate distribution of the firm's customer base, $\mathbf{E}_{\hat{\pi}}$ as the expectation taken w.r.t the sampling distribution of $\hat{\pi}_S$, and $\mathbf{E}$ as the expectation taken w.r.t both uncertainty.

We establish three key properties: For any sufficiently large initial set $\mathcal{D}$ and any augmented set $S = \mathcal{D} \cup S_k \subseteq \mathcal{I}$ with $k > 0$ and $|S_k| = k$, the objective function f satisfies the following conditions:

1. $f(S)$ is monotonically non-decreasing with respect to set inclusion.

2. $f(S)$ is submodular, i.e., the marginal benefit of adding an element diminishes as the set grows.

3. The expected profit loss criterion:

$$\ell_S(\mathbf{X}_i) = \mathbb{E}_{\mathcal{D}}\left[|\tau(\mathbf{X}_i) - c(\mathbf{X}_i)| \cdot \mathbf{1}\{\hat{\pi}_S(\mathbf{X}_i) \neq \pi^*(\mathbf{X}_i)\}\right]$$

   identifies the elements that yield the maximum marginal improvement in the objective function f.

Consequently, iteratively selecting customers with the largest $\ell_S(\mathbf{X}_i)$ after initializing with $\mathcal{D}$ constitutes a greedy algorithm. By the classic result of Nemhauser et al. (1978), we have

$$f(\mathcal{D} \cup S_k^g) \geq (1 - \frac{1}{e}) \cdot f(\mathcal{D} \cup S_k^*),$$

where S_g^k denotes the k points selected by the greedy algorithm and S_k^* represents the optimal set of k points such that the targeting policy $\hat{\pi}_S$ estimated from it maximizes the firm's expected profit when deployed on future customers with the same distribution as $\mathcal{I}$.

Monotonicity

We first show that the objective function $f(S)$ is monotonically non-decreasing as we gather more samples. Let us define

$$\begin{aligned}
L(S) &= f^* - f(S) \\
&= \mathbb{E}\left[Y_i(0) + (\tau(\mathbf{X}_i) - c(\mathbf{X}_i)) \cdot \pi^*(\mathbf{X}_i)\right] - \mathbb{E}\left[Y_i(0) + (\tau(\mathbf{X}_i) - c(\mathbf{X}_i)) \cdot \hat{\pi}_S(\mathbf{X}_i)\right] \\
&= \mathbf{E}\left[|\tau(\mathbf{X}_i) - c(\mathbf{X}_i)| \cdot \mathbf{1}\{\hat{\pi}_S(\mathbf{X}_i) \neq \pi^*(\mathbf{X}_i)\}\right] \\
&= \mathbf{E}_{\mathbf{X}}\left[|\tau(\mathbf{X}_i) - c(\mathbf{X}_i)| \cdot \Pr(\mathbf{1}\{\hat{\pi}_S(\mathbf{X}_i) \neq \pi^*(\mathbf{X}_i)\})\right].
\end{aligned}$$

By the symmetric property of the error-density, we have that

$$\Pr(\mathbf{1}\{\hat{\pi}_S(x) \neq \pi^*(x)\}) = F_Z\left(\frac{-|\tau(x) - c(x)|}{\sigma(x)}\right) \quad \forall x \in \mathbf{X},$$

where $Z = \frac{(\hat{\tau}(x) - \tau(x))}{\sigma(x)}$. By the variance shrinking property, $\forall x \in \mathbf{X}$, $\sigma_S^2(x) \geq \sigma_T^2(x)$ if $S \subseteq T$. Since $F_Z\left(\frac{-|\tau(x)-c(x)|}{\sigma(x)}\right)$ is weakly increasing in $\sigma(x)$, the mistargeting probability is weakly increasing in $\sigma(x)$, implying that

$$|\tau(x) - c(x)| \cdot \Pr(\mathbf{1}\{\hat{\pi}_S(x) \neq \pi^*(x)\}) \geq |\tau(x) - c(x)| \cdot \Pr(\mathbf{1}\{\hat{\pi}_T(x) \neq \pi^*(x)\})$$

for all $x \in \mathbf{X}$. Taking the expectation over the covariate distribution, we have that

$$L(S) \geq L(T),$$

and since $f(S) = f^* - L(S)$, we therefore have

$$f(S) \leq f(T).$$

Therefore the objective function is monotonically non-decreasing.

Submodularity

Next, we show that the objective function $f(S)$ is submodular, i.e.,

$$f(S \cup \{j\}) - f(S) \geq f(T \cup \{j\}) - f(T)$$

for any $S \subseteq T \subset \mathcal{I}$ and $j \notin T$.

We start by showing that $\forall x \in \mathbf{X}$, the pointwise loss improvement

$$\begin{aligned}
\ell_S(x) &= \mathbb{E}_{\hat{\pi}}\left[|\tau(x) - c(x)| \cdot \mathbf{1}\{\hat{\pi}_S(x) \neq \pi^*(x)\}\right] \\
&= |\tau(x) - c(x)| \cdot \Pr(\mathbf{1}\{\hat{\pi}_S(x) \neq \pi^*(x)\}) \\
&= |\tau(x) - c(x)| \cdot F_Z\left(\frac{-|\tau(x) - c(x)|}{\sigma_S(x)}\right) \\
&= |\tau(x) - c(x)| \cdot F_Z\left(\frac{-|\tau(x) - c(x)|}{\sigma(x)} \cdot \sqrt{\gamma_{|S|}(x)}\right)
\end{aligned}$$

satisfies

$$\ell_S(x) - \ell_{S \cup \{j\}}(x) \geq \ell_T(x) - \ell_{T \cup \{j\}}(x).$$

Since $F_Z(\cdot)$ is a continuous non-decreasing function, by mean-value theorem,

$$\begin{aligned}
&F_Z\left(\frac{-|\tau(x) - c(x)|}{\sigma(x)} \cdot \sqrt{\gamma_{|S|}(x)}\right) - F_Z\left(\frac{-|\tau(x) - c(x)|}{\sigma(x)} \cdot \sqrt{\gamma_{|S+1|}(x)}\right) = \\
&f_Z(u_{S,x}) \cdot \underbrace{\frac{|\tau(x) - c(x)|}{\sigma(x)}}_{d(x)} \cdot \left(\sqrt{\gamma_{|S+1|}(x)} - \sqrt{\gamma_{|S|}(x)}\right)
\end{aligned}$$

for the interval where $u_{S,x} \in \left(-d(x) \cdot \sqrt{\gamma_{|S+1|}(x)}, -d(x) \cdot \sqrt{\gamma_{|S|}(x)}\right)$.

By the log-concavity of the error-density function, we have that

$$f_Z(u_{S,x}) \geq f_Z(u_T)$$

as $u_{S,x} > u_{T,x}$. Furthermore, due to the variance-shrinking property and the concavity of square-root function, we have that

$$\sqrt{\gamma_{|S+1|}(x)} - \sqrt{\gamma_{|S|}(x)} \geq \sqrt{\gamma_{|T+1|}(x)} - \sqrt{\gamma_{|T|}(x)}.$$

Putting all together, for any $x \in \mathbf{X}$, we have that

$$\ell_S(x) - \ell_{S \cup \{j\}}(x) \geq \ell_T(x) - \ell_{T \cup \{j\}}(x).$$

Taking expectation over the covariate distribution, we have

$$L(S) - L(S \cup \{j\}) \geq L(T) - L(T \cup \{j\}),$$

implying that

$$f(S \cup \{j\}) - f(S) \geq f(T \cup \{j\}) - f(T).$$

Therefore f is submodular.

Greediness of expected profit loss sampling

We now establish the greediness of the proposed algorithm. In particular, we show that given a sufficiently large initial set $\mathcal{D}$, for any $S = \mathcal{D} \cup S_k$ with $k > 0$ and $|S_k| = k$, the element producing the largest marginal improvement $f(S \cup \{j\}) - f(S)$ is identical to the one with the largest expected profit loss, i.e.,

$$\arg\max_j f(S \cup \{j\}) - f(S) = \arg\max_j \ell_S(\mathbf{X}_j).$$

We first show that for distinct indices $i \neq j$, the difference between the point-wise marginal improvement of point i and that of point j is asymptotically negligible, i.e.,

$$\ell_S(\mathbf{X}_i) - \ell_{S\cup\{j\}}(\mathbf{X}_i) = \begin{cases} (\ell_S(\mathbf{X}_j) - \ell_{S\cup\{j\}}(\mathbf{X}_j)) \cdot (1 + O(|S|^{-\alpha})) & \text{if } \mathbf{X}_i \in B(\mathbf{X}_j, h_{|S|}) \\ 0 & \text{otherwise.} \end{cases}$$

If $\mathbf{X}_i \notin B(\mathbf{X}_j, h_{|S|})$, since i is out of the influence region of $\mathbf{X}_j$, $\ell_S(\mathbf{X}_i) - \ell_{S\cup\{j\}}(\mathbf{X}_i) = 0$. If $\mathbf{X}_i \in B(\mathbf{X}_j, h_{|S|})$, since

$$\ell_S(x) - \ell_{S\cup\{j\}}(x) = |\tau(x) - c(x)| \cdot f_Z(u_{S,x}) \cdot \underbrace{\frac{|\tau(x) - c(x)|}{\sigma(x)}}_{d(x)} \cdot (\sqrt{\gamma_{|S+1|}(x)} - \sqrt{\gamma_{|S|}(x)}),$$

we have that

$$\frac{\ell_S(\mathbf{X}_i) - \ell_{S\cup\{j\}}(\mathbf{X}_i)}{(\ell_S(\mathbf{X}_j) - \ell_{S\cup\{j\}}(\mathbf{X}_j))} = \frac{|\tau(\mathbf{X}_i) - c(\mathbf{X}_i)|}{|\tau(\mathbf{X}_j) - c(\mathbf{X}_j)|} \cdot \frac{f_Z(u_{S,i})}{f_Z(u_{S,j})} \cdot \frac{d(\mathbf{X}_i)}{d(\mathbf{X}_j)} \cdot \frac{\sqrt{\gamma_{|S+1|}(\mathbf{X}_i)} - \sqrt{\gamma_{|S|}(\mathbf{X}_i)}}{\sqrt{\gamma_{|S+1|}(\mathbf{X}_j)} - \sqrt{\gamma_{|S|}(\mathbf{X}_j)}}. \tag{W-1}$$

Since the functions $f_Z(\cdot)$, $\sigma(\cdot)$, and $\gamma_n(\cdot)$ are continuously differentiable on their respective domains, they satisfy the local Lipschitz condition. Specifically, for any compact ball $B \subset \mathcal{X}$, there exists a constant $L_B > 0$ such that each function is L_B-Lipschitz on B. Furthermore, given that

$\tau(\cdot)$ and $c(\cdot)$ are both L-Lipschitz on $\mathcal{X}$, the function $d(\cdot) = \frac{|\tau(\cdot)-c(\cdot)|}{\sigma(\cdot)}$ and the term $\sqrt{\gamma_{|S|+1}(\cdot)} - \sqrt{\gamma_{|S|}(\cdot)}$ also satisfy the local Lipschitz condition.

Therefore, since $\mathbf{X}_i \in B(\mathbf{X}_j, h_{|S|})$ and the interval length of $u_{S,x}$: $d(x)\cdot(\sqrt{\gamma_{|S+1|}(x)} - \sqrt{\gamma_{|S|}(x)}) = O(|S|^{-\frac{1+\alpha d}{2}}) < O(h_{|S|})$, we have

$$
\begin{aligned}
|u_{S,i} - u_{S,j}| &\leq \left| d(\mathbf{X}_i)\cdot\sqrt{\gamma_{|S|+1}(\mathbf{X}_j)} - d(\mathbf{X}_j)\sqrt{\gamma_{|S|}(\mathbf{X}_j)} \right| \\
&\leq \underbrace{\left| d(\mathbf{X}_i)\cdot(\sqrt{\gamma_{|S|+1}(\mathbf{X}_i)} - d(\mathbf{X}_i)\sqrt{\gamma_{|S|}(\mathbf{X}_i)} \right|}_{\text{interval length of } u_{s,i} = O(|S|^{-\frac{1+\alpha d}{2}})} + \left| d(\mathbf{X}_i)\cdot(\sqrt{\gamma_{|S|}(\mathbf{X}_i)} - d(\mathbf{X}_j)\sqrt{\gamma_{|S|}(\mathbf{X}_j)} \right| \\
&\leq O(|S|^{-\frac{1+\alpha d}{2}}) + O(h_{|S|}) \\
&\leq O(h_{|S|}),
\end{aligned}
$$

where the second inequality comes from the triangular inequality and the third comes from the local Lipschitz condition of $d(\cdot)$ and $\gamma(\cdot)$. Thus, by the local Lipschitz condition, we have that

$$
\begin{aligned}
|\tau(\mathbf{X}_i) - c(\mathbf{X}_i)| &= |\tau(\mathbf{X}_j) - c(\mathbf{X}_j)| + O(h_{|S|}) \\
f_Z(u_{S,i}) &= f_Z(u_{S,j}) + O(h_{|S|}) \\
d(\mathbf{X}_i) &= d(\mathbf{X}_j) + O(h_{|S|}) \\
\sqrt{\gamma_{|S+1|}(\mathbf{X}_i)} - \sqrt{\gamma_{|S|}(\mathbf{X}_i)} &= \sqrt{\gamma_{|S+1|}(\mathbf{X}_j)} - \sqrt{\gamma_{|S|}(\mathbf{X}_j)} + O(h_{|S|}).
\end{aligned}
$$

Plugging into Equation (W-1), we have that

$$
\frac{\ell_S(\mathbf{X}_i) - \ell_{S\cup\{j\}}(\mathbf{X}_i)}{(\ell_S(\mathbf{X}_j) - \ell_{S\cup\{j\}}(\mathbf{X}_j))} = 1 + O(h_{|S|}).
$$

Taking expectation over the covariate distribution, we have that

$$
\begin{aligned}
L(S) - L(S\cup\{j\}) &= \mathbf{E}_{\mathbf{X}}[\ell_S(\mathbf{X}_i) - \ell_{S\cup\{j\}}(\mathbf{X}_i)] \\
&\propto (\ell_S(\mathbf{X}_j) - \ell_{S\cup\{j\}}(\mathbf{X}_j))\cdot(1 + O(h_{|S|})) \qquad \text{(W-2)} \\
&\approx \ell_S(\mathbf{X}_j) - \ell_{S\cup\{j\}}(\mathbf{X}_j)
\end{aligned}
$$

since $|S| = |\mathcal{D}| + k$ is sufficiently large.

We now show that for any $i, j \notin S$, if $\ell_S(\mathbf{X}_j) \geq \ell_S(\mathbf{X}_i)$, $f(S\cup\{j\}) - f(S) \geq f(S\cup\{i\}) - f(S)$. By the variance-shrinking property, since $\ell_S(\mathbf{X}_j) \geq \ell_{S\cup\{j\}}(\mathbf{X}_j)$, there exists a $\beta_j \in (0,1)$ such that

$$
\ell_{S\cup\{j\}}(\mathbf{X}_j) \leq \beta\ell_S(\mathbf{X}_j).
$$

Consequently, if $\ell_S(\mathbf{X}_j) \geq \ell_S(\mathbf{X}_i)$, let $\beta = \max\{\beta_i, \beta_j\}$, we obtain that

$$\begin{aligned}
(L(S) - L(S \cup \{j\})) - (L(S) - L(S \cup \{i\})) &\approx (\ell_S(\mathbf{X}_j) - \ell_{S\cup\{j\}}(\mathbf{X}_j)) - (\ell_S(\mathbf{X}_i) - \ell_{S\cup\{i\}}(\mathbf{X}_i)) \\
&= (\ell_S(\mathbf{X}_j) - \ell_S(\mathbf{X}_i)) - (\ell_{S\cup\{j\}}(\mathbf{X}_j) - \ell_{S\cup\{i\}}(\mathbf{X}_i)) \\
&\geq (\ell_S(\mathbf{X}_j) - \ell_S(\mathbf{X}_i)) - \beta(\ell_S(\mathbf{X}_j) - \ell_S(\mathbf{X}_i)) \\
&= (1-\beta)(\ell_S(\mathbf{X}_j) - \ell_S(\mathbf{X}_i)) \\
&\geq 0.
\end{aligned}$$

Therefore, if $\ell_S(\mathbf{X}_j) \geq \ell_S(\mathbf{X}_i)$,

$$L(S) - L(S \cup \{j\}) \geq L(S) - L(S \cup \{i\}),$$

implying that

$$f(S \cup \{j\}) - f(S) \geq f(S \cup \{i\}) - f(S).$$

Thus, for any $S = \mathcal{D} \cup S_k$ with $k > 0$,

$$\arg\max_j f(S \cup \{j\}) - f(S) = \arg\max_j \ell_S(\mathbf{X}_j).$$

Conclusion

Given a sufficiently large initial sample $\mathcal{D}$, since for any $S = \mathcal{D} \cup S_k \in \mathcal{I}$ with $k > 0$ and $|S_k| = k$, we have established that: (i) the objective function $f(S)$ is monotone, (ii) $f(S)$ is submodular, and (iii) the expected profit loss function

$$\ell_S(\mathbf{X}_i) = \mathbb{E}_{\hat{\pi}}\left[|\tau(\mathbf{X}_i) - c(\mathbf{X}_i)| \cdot \mathbf{1}\{\hat{\pi}_S(\mathbf{X}_i) \neq \pi^*(\mathbf{X}_i)\}\right]$$

identifies the customers who yield the largest marginal improvement in f. By the classical result of Nemhauser et al. (1978), the greedy algorithm that iteratively selects customers with the highest expected profit loss $\ell_S(\mathbf{X}_i)$ after initializing with $\mathcal{D}$ satisfies the following approximation guarantee

$$\begin{aligned}
f(\mathcal{D} \cup S_k^g) &\geq f(\mathcal{D}) + (1 - \frac{1}{e}) \cdot (f(\mathcal{D} \cup S_k^*) - f(\mathcal{D})) \\
&\geq (1 - \frac{1}{e}) f(\mathcal{D} \cup S_k^*),
\end{aligned}$$

where S_k^g denotes the greedy solution and S_k^* the optimal sample of size k. This concludes the proof.

□

C. Further Details of Simulation Studies

C.1. Implementation Details

C.1.1. Default (A/B Test with Random Sampling)

For each experimental size $N_S \in \{1k, 5k, 10k, 20k, 30k\}$, we randomly sample N_S customers from the customer base $\mathcal{I}$ and assign them randomly to two treatment conditions with a probability of 0.5. For CATE estimation, we construct a Causal Forest model (Wager and Athey 2018) implemented using the `econML` package in Python. This model consists of 100 trees with a maximum depth of 10 to prevent overfitting.

C.1.2. Uncertainty Sampling

For each experimental size $N_S \in \{1k, 5k, 10k, 20k, 30k\}$, we follow a similar procedure as Waisman et al. (2024) by using equal-sized batches throughout the experiment. In particular, we partition the full sample into $\frac{N_S}{200}$ batches, each containing 200 customers. Additionally, we evaluate alternative sample allocation schemes (constant batch size and decreasing batch size over 10 batches) to assess their impact on performance. Customers within each batch b are randomly assigned to the two treatment conditions with a probability of 0.5.

For uncertainty estimation, as well as for the final CATE estimation, we utilize a Causal Forest model implemented using the `econML` package in Python. This model consists of 100 trees with maximum depths not exceeding 10. We estimate the uncertainty associates with the CATE estimates $\sigma(\hat{\tau}_{S^{b-1}}(x))$ by computing the standard deviation across the predictions generated by each tree.

C.1.3. Kato et al. (2024)

For each experimental size $N_S \in \{1k, 5k, 10k, 20k, 30k\}$, we follow a similar procedure as Waisman et al. (2024) by using equal-sized batches throughout the experiment. Specifically, we partition the full sample into $\frac{N_S}{200}$ batches, and randomly sample 200 customers who have not been sampled in previous batches from the customer base $\mathcal{I}$ in each batch b. Additionally, we evaluate alternative sample allocation schemes (constant batch size and decreasing batch size over 10 batches) to assess their impact on performance.

For customers in the first batch $b = 1$, we assign them randomly to the two treatment conditions with a probability of 0.5. For customers subsequent batches, we assign them to the two treatment arms based on the following rule:

$$P_{S^{b-1}}(W_i = 0|\mathbf{X}_i = x) = \frac{\sigma^0_{S^{b-1}}(x)}{\sigma^0_{S^{b-1}}(x) + \sigma^1_{S^{b-1}}(x)}$$
$$P_{S^{b-1}}(W_i = 1|\mathbf{X}_i = x) = \frac{\sigma^1_{S^{b-1}}(x)}{\sigma^0_{S^{b-1}}(x) + \sigma^1_{S^{b-1}}(x)}$$

where $\sigma^w_{S^{b-1}}(x)$ denotes the standard deviation of the potential outcomes $Y_i(W_i = w)$ estimated from the previous $b-1$ batches. In particular, we estimate customer's response function for the two treatment conditions $(\mathbf{E}[Y_i(0)|\mathbf{X}_i], \mathbf{E}[Y_i(1)|\mathbf{X}_i])$ using two Random Forest models. These Random Forest models are implemented using the `sklearn` package in Python, each consisting of 100 trees with maximum depths not exceeding 10. We estimate $(\sigma^0_{S^{b-1}}(x), \sigma^1_{S^{b-1}}(x))$ by computing the standard deviation across the predictions generated by each tree.

We slightly modify the decision estimation phase from the original paper to ensure comparability with our approach. Specifically, we derive the targeting decisions $\hat{\pi}(\mathbf{X}_i)$ using the CATE predictions $\hat{\tau}(\mathbf{X}_i)$ generated by the CATE model, rather than directly estimating them with a policy learning model (e.g. Athey and Wager 2021). This adjustment allows us to eliminate potential differences in targeting performance that may arise from different estimation strategies.[28]

For CATE estimation, we construct a Causal Forest model implemented using the `econML` package in Python, consisting of 100 trees with maximum depths not exceeding 10. Note that the Causal Forest model in `econML` is designed to solve the local moment equation:

$$\mathbf{E}[Y_i - \tau(x) \cdot W_i - B(x)|\mathbf{X}_i = x] = 0$$

where $B(x) = \mathbf{E}[Y_i|\mathbf{X}_i = x]$. Therefore, we account for the adaptive nature of the experimental data by substracting the propensity score $P_{S^{b-1}}(W_i = 1|\mathbf{X}_i = x) = \frac{\sigma^1_{S^{b-1}}(x)}{\sigma^0_{S^{b-1}}(x)+\sigma^1_{S^{b-1}}(x)}$ from the actual treatment assignment W_i based on Robinson's Decomposition (Robinson 1988):

$$Y_i - B(\mathbf{X}_i) = \tau(\mathbf{X}_i)(W_i - e(\mathbf{X}_i)) + \varepsilon_i$$

[28]Note that the treatment assignment ratio proposed in Kato et al. (2024) remains unaffected by the estimation strategy.

C.1.4. Proposed Approach with Multiple Stages

For each experimental size $N_S \in \{1k, 5k, 10k, 20k, 30k\}$, we follow a similar procedure as Waisman et al. (2024) by using equal-sized batches throughout the experiment. In particular, we partition the full sample into $\frac{N_S}{200}$ batches, each containing 200 customers. Additionally, we evaluate alternative sample allocation schemes (constant batch size and decreasing batch size over 10 batches) to assess their impact on performance. Customers within each batch b are randomly assigned to the two treatment conditions with a probability of 0.5.

For EPL estimation, as well as for the final CATE estimation, we utilize a Causal Forest model implemented using the `econML` package in Python. This model consists of 100 trees with maximum depths not exceeding 10.

C.1.5. Proposed Approach with Two Stages

For each experimental size $N_S \in \{1k, 5k, 10k, 20k, 30k\}$, we consider three different proportions of customers r to sample in the first stage ($r \in \{0.5, 0.7, 0.9\}$) and follow a two-stage sampling approach:

1. In the first stage, we randomly sample $r \cdot N_S$ customers from the customer base $\mathcal{I}$.

2. In the second stage, we select the remaining $(1 - r) \cdot N_S$ customers who have the highest EPL estimated from the first stage.

3. Customers sampled in both stages are randomly assigned to the two treatment conditions with a probability of 0.5.

For EPL and final CATE estimation, we employ a Causal Forest model implemented using the `econML` package in Python. This model comprises 100 trees with maximum depths not exceeding 10.

C.2. Additional Results for Alternative Sample Allocation Schemes

In this appendix, we present additional results examining alternative sample allocation schemes with the same CATE distribution and intervention costs as in Section 5. Specifically, for the multi-stage design, we evaluate two sample allocation approaches: constant batch size and decreasing batch size over 10 batches. In addition, we explore three different configurations of the two-stage design by varying the proportions of customers r to sample in the first stage ($r \in \{0.5, 0.7, 0.9\}$) and compare their performance with the multi-stage design.

C.2.1. Alternative Sample Allocation Schemes for Multi-Stage Design

We first examine the performance of the multi-stage design of our approach under two alternative sample allocation schemes: a constant batch size allocation (dividing the experimental budget equally across 10 batches) with a decreasing batch size allocation (where earlier batches receive more samples).

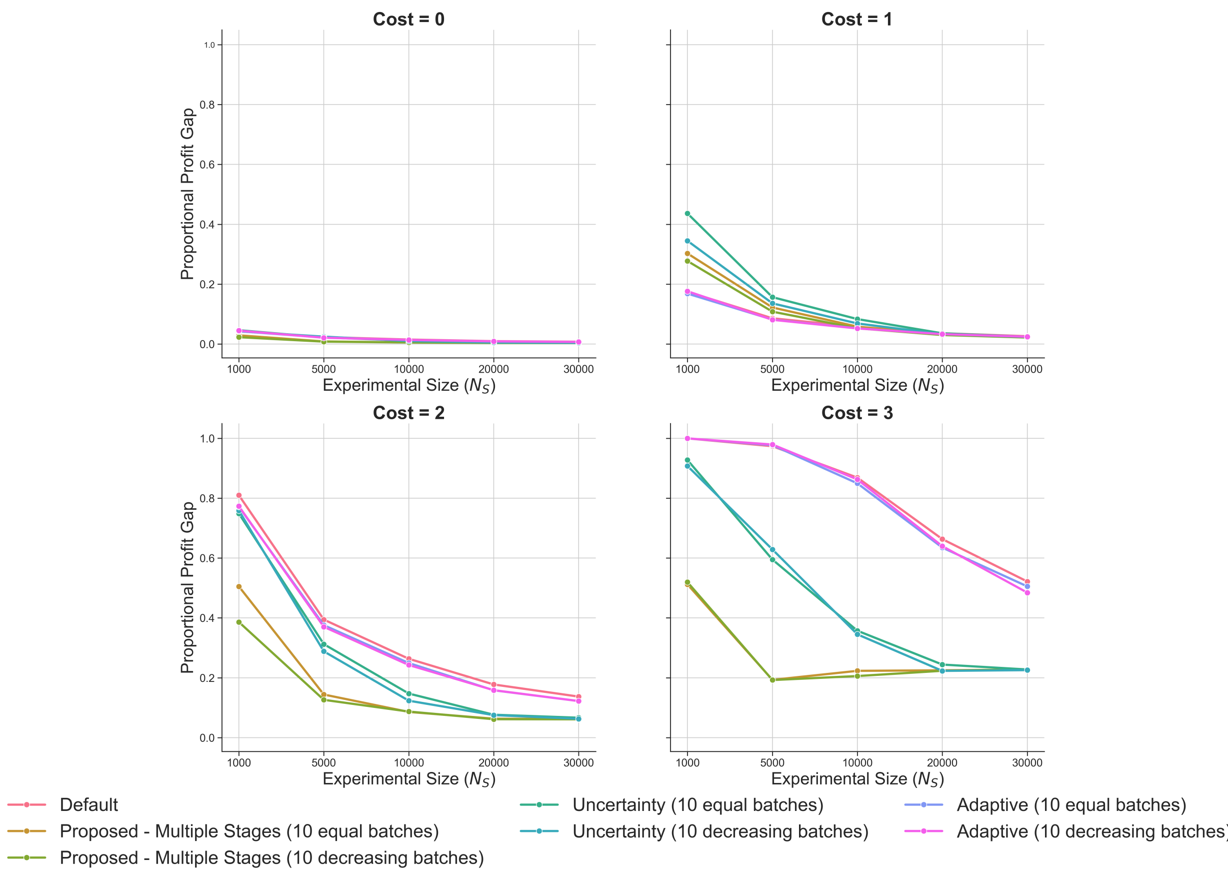


Figure W-1: Proportional Profit Gaps of Different Experimental Designs under Alternative Sample Allocation Schemes
We report the average value of the proportional profit gap across 100 replications. Each line corresponds to an experimental approach.

Figure W-1 shows the proportional profit gaps of the targeting policies learned by different experimental designs under different sample allocation schemes across different intervention costs ($c \in \{0, 1, 2, 3\}$) and experimental sizes ($N_S \in \{1k, 5k, 10k, 20k, 30k\}$) respectively. The results are qualitatively similar to the one with the baseline fixed batch size of 200. In particular, our approach consistently outperforms the three benchmarks when the number of customers around the decision threshold is limited (i.e., $c \in \{0, 2, 3\}$), especially at smaller sample sizes. Notably, the performance of our approach remains similar across both alternative allocation schemes (constant and decreasing batch sizes), with decreasing batch sizes showing slight advantages when experimental budgets are particularly small ($N_S = 1k$), indicating that the specific allocation strategy has minimal impact on the effectiveness of our approach.

C.2.2. Alternative Configurations for Two-Stage Design

We now explore the performance of the simplified two-stage design of our approach under three different proportions of customers r to sample in the first stage: $r \in \{0.5, 0.7, 0.9\}$.

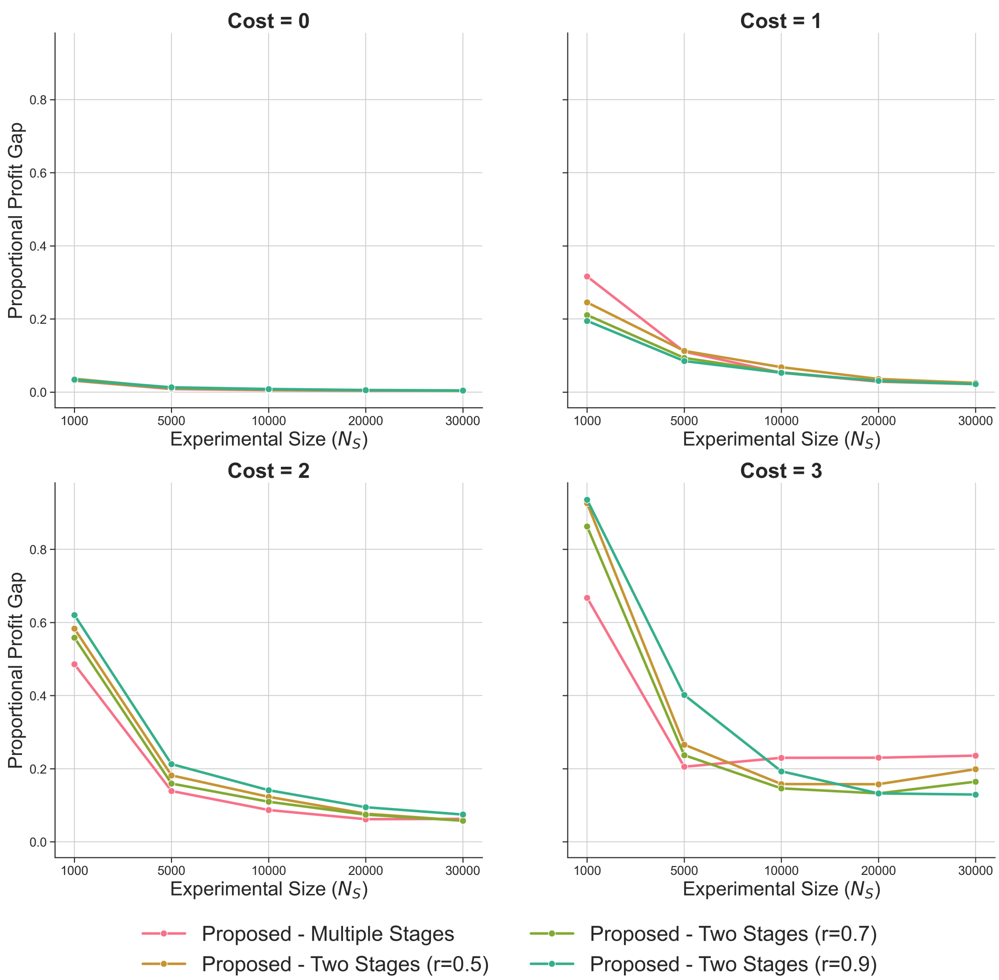


Figure W-2: Proportional Profit Gaps of the Two-Stage Design under Alternative Configurations

We report the average value of the proportional profit gap across 100 replications. Each line corresponds to an experimental approach.

Figure W-2 shows the proportional profit gaps of the targeting policies learned by a multi-stage design and three two-stage design of our approach under different configurations ($r \in \{0.5, 0.7, 0.9\}$) across different intervention costs ($c \in \{0, 1, 2, 3\}$) and experimental sizes ($N_S \in \{1k, 5k, 10k, 20k, 30k\}$) respectively. The results indicate that the two-stage design generally performs comparably to the multiple-stage version. Moreover, when the cost is near the mode and the sample size is small (e.g., $N_S = 1000$), a two-stage design with a higher proportion of customers sampled in the first stage (i.e., $r = 0.9$) demonstrates the best performance among the four variants of our proposed approach. This is because in such scenarios, more accurate EPL estimates are crucial to effectively sample the consequential customers as intensively as random sampling does. Hence, a greater proportion of customers sampled in the first stage becomes necessary when dealing with smaller sample sizes.

C.3. Additional Results for Alternative CATE distributions

In this appendix, we present additional robustness results examining different CATE distributions. Specifically, we consider two additional distributions: a bimodal distribution with two equal segments and a bimodal distribution with two unequal segments. We test these two additional CATE distributions using the three sample allocation schemes for the multi-stage design (fixed batch size of 200, constant batch size across 10 batches, and decreasing batch size across 10 batches) as well as the three configurations of the two-stage design ($r \in \{0.5, 0.7, 0.9\}$).

C.3.1. Bimodal Distribution with Two Equal Segments

We generate a customer base $\mathcal{I}$ with a bimodal CATE distribution featuring two equal segments according to the following data generating process:

$$Y_i = \tau(X_i) * W_i + X_{i4} * X_{i5} + e_i, \quad e_i \sim \mathcal{N}(0, 1)$$
$$\tau(X_i) = X_{i1} * X_{i2} + X_{i3} * (1 - X_{i2})$$

where

$$X_{i1} \sim \mathcal{N}(2, 1.5)$$
$$X_{i3} \sim \mathcal{N}(-4, 1.5)$$
$$X_{i5} \sim \mathcal{N}(0, 1)$$
$$X_{ij} \sim Bernoulli(0.5), \quad j \in \{2, 4\}$$

are identically and independently distributed.

In this scenario, we consider three different intervention costs $c \in \{-4, -1, 2\}$ where $c = -4$ and $c = 2$ corresponds to peak of the distribution, and $c = -1$ represents a valley in the CATE distribution. Figure W-3 visualizes the relationship between the CATE distribution and the intervention costs.

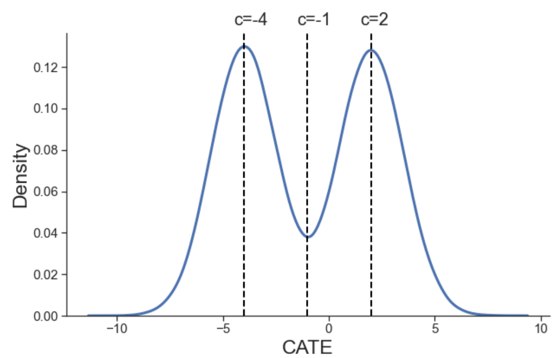


Figure W-3: CATE Distribution and Intervention Costs: Bimodal Distribution with Two Equal Segments

Each dashed line corresponds to a different intervention cost c. $c \in \{-4, 2\}$ aligns with the mode of the CATE distribution. $c = -1$ represents the valley of the CATE distribution.

Figure W-4 and Figure W-5 shows the proportional profit gaps and profit gaps of the targeting policies learned by different experimental designs across different intervention costs ($c \in \{-4, -1, 2\}$) and experimental sizes ($N_S \in \{5k, 10k, 20k, 30k\}$) respectively.[29] The results are qualitatively similar to the one with normally distributed CATEs. In particular, our approach consistently outperforms the three benchmarks when the number of customers around the decision threshold is limited (i.e., $c = -1$). Conversely, when the decision threshold aligns with the mode of the distribution (i.e., $c \in \{-4, 2\}$), the improvement against the default and approach and Kato et al. (2024) is subtle. [30]

Figure W-6 further shows the proportional profit gaps of the targeting policies learned by different experimental designs under different sample allocation schemes. The results are qualitatively similar to the one with the baseline fixed batch size of 200. In particular, our approach consistently outperforms the three benchmarks when the number of customers around the decision threshold is limited (i.e., $c = -1$). While the two allocation schemes generally produce

[29]When $c = -1$, due to the scarcity of consequential customers, our approach requires a larger sample size to identify sufficient consequential customers. Therefore, we omit the analysis with $N_S = 1k$ for this case.

[30]The difference in the proportional profit gap between $c = -4$ and $c = 2$ is primarily due to the disparity in their denominators. In particular, since $c = -4$ represents a less costly intervention, the incremental profit generated by the optimal policy is greater compared to $c = 2$. As a result, despite having similar numerators, this leads to asymmetry in the proportional profit gaps between the two scenarios.

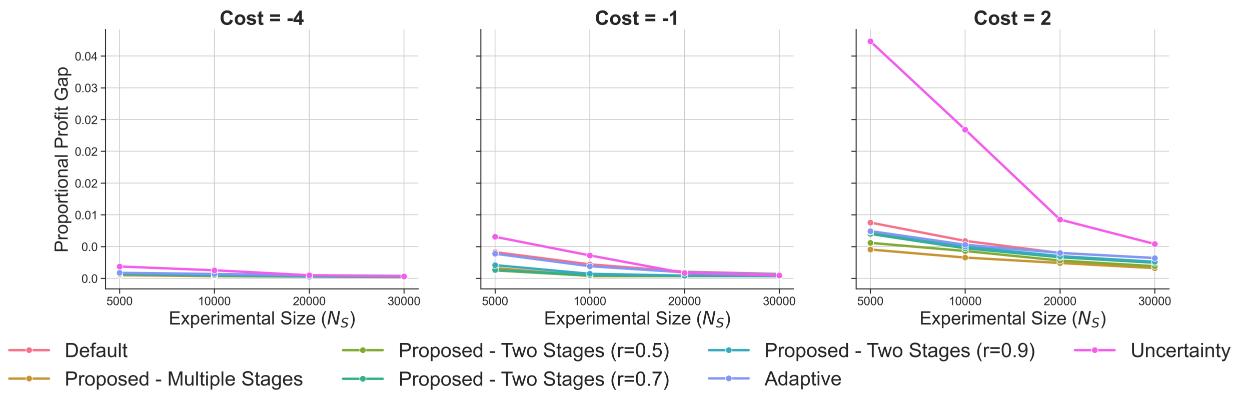


Figure W-4: Proportional Profit Gaps of Different Experimental Designs: Bimodal Distribution with Two Equal Segments

We report the average value of the proportional profit gap across 100 replications. Each line corresponds to an experimental approach.

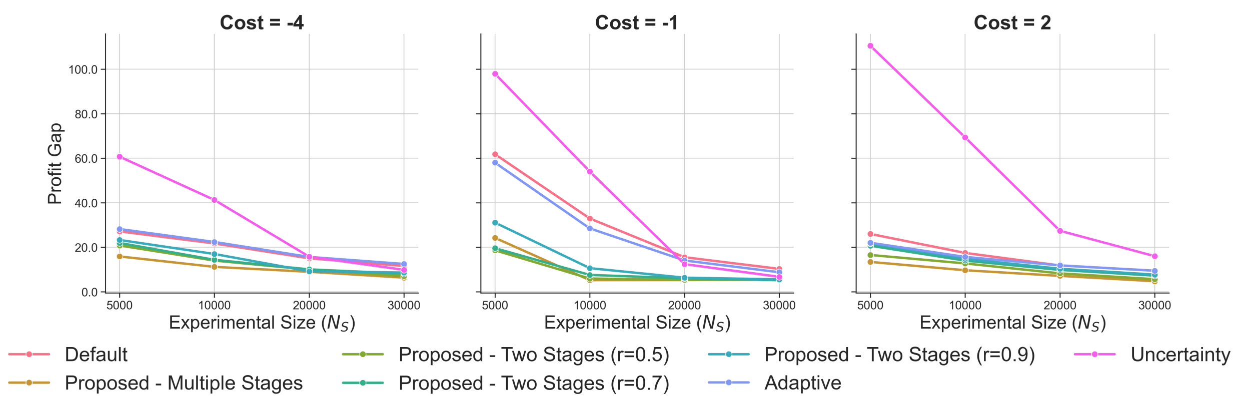


Figure W-5: Profit Gaps of Different Experimental Designs: Bimodal Distribution with Two Equal Segments

We report the average value of the profit gap across 100 replications. Each line corresponds to an experimental approach.

comparable performance, the decreasing batch size shows slight advantages when sample size is small, especially when the intervention cost aligns with the mode $c = 2$).

Figure W-7 displays the CATE distributions of the customers sampled by various approaches across different intervention costs. The results underscore the effectiveness of our approach in identifying and intensively sampling consequential customers.

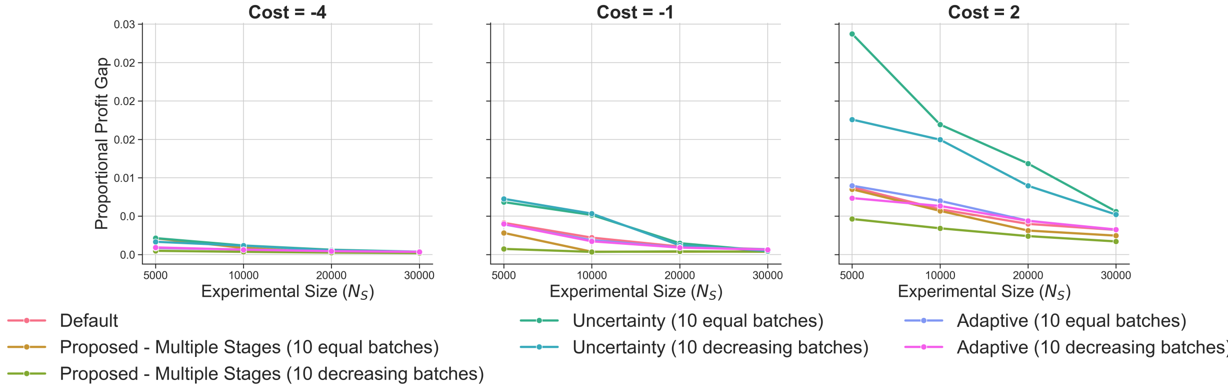


Figure W-6: Proportional Profit Gaps of Different Experimental Designs under Alternative Sample Allocation Schemes

We report the average value of the proportional profit gap across 100 replications. Each line corresponds to an experimental approach.

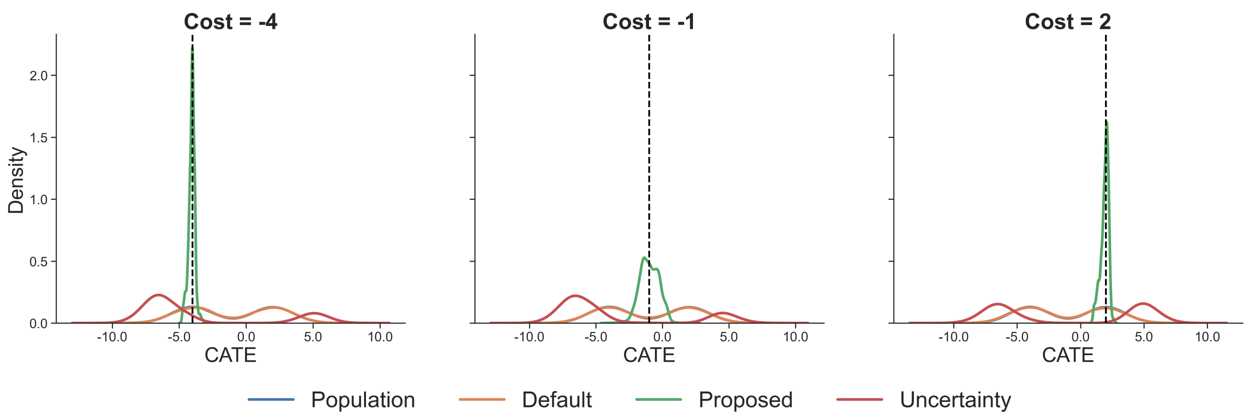


Figure W-7: CATE Distributions of Customers Sampled by Different Approaches: Bimodal Distribution with Two Equal Segments

Each line corresponds to the CATE distribution of the customers sampled by different approaches. The dashed line represents the intervention cost, which is also the decision threshold. The difference between $c = -4$ and $c = 2$ is mainly driven by the difference in the denominator. In particular, since $c = 2$ exhibits greater cost, the profit of the optimal policy of $c = 2$ is smaller than $c = -4$, leading to a smaller denominator in the proportional profit gap formula. Therefore, the evaluation metric

C.3.2. Bimodal Distribution with Two Unequal Segments

We generate a customer base $\mathcal{I}$ with a bimodal CATE distribution featuring two unequal segments according to the following data generating process:

$$Y_i = \tau(X_i) * W_i + X_{i4} * X_{i5} + e_i, \quad e_i \sim \mathcal{N}(0,1)$$
$$\tau(X_i) = X_{i1} * X_{i2} + X_{i3} * (1 - X_{i2})$$

where

$$X_{i1} \sim \mathcal{N}(3,1)$$
$$X_{i3} \sim \mathcal{N}(-2,1)$$
$$X_{i5} \sim \mathcal{N}(0,1)$$
$$X_{ij} \sim Bernoulli(0.5), \quad j \in \{2,4\}$$

are identically and independently distributed.

In this scenario, we examine three different intervention costs $c \in \{-2, 1, 3\}$ where $c = -2$ is at the lager peak, $c = 3$ is at the smaller peak, and $c = 1$ is at the valley of the CATE distribution. Figure W-8 illustrates the relationship between the CATE distribution and the intervention costs.

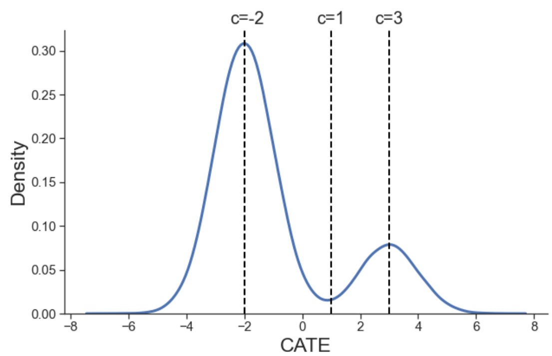


Figure W-8: CATE Distribution and Intervention Costs: Bimodal Distribution with Two Unequal Segments

Each dashed line corresponds to a different intervention cost c. $c = -2$ aligns with the larger peak and $c = 3$ aligns with the smaller peak of the CATE distribution. $c = 1$ represents the valley of the CATE distribution.

Figure W-9 shows the proportional profit gaps of the targeting policies learned by different experimental designs across different intervention costs ($c \in \{-2, 1, 3\}$) and experimental sizes ($N_S \in \{5k, 10k, 20k, 30k\}$).[31]

As shown in the graph, our approach generally surpasses the three benchmarks when the cost is at the valley and the smaller peak of the CATE distribution.[32] However, when the sample size is small (i.e., $N_S = 5000$), it is crucial for the firm to carefully select the appropriate deign. Specifically, when the decision threshold is at the smaller peak, the firm should avoid a two-stage design with a limited first-stage sample, as the small initial sample size can lead to substantial EPL esti-

[31]When $c = 1$, due to the scarcity of consequential customers, our approach requires a larger sample size to identify sufficient consequential customers. Therefore, we omit the analysis with $N_S = 1k$ for this case.

[32]When the decision threshold aligns with the valley of the distribution, our approach still outperforms the benchmarks as expected, albeit with smaller magnitude. This reduced improvement occurs because most customers are inconsequential in this case — their prediction errors have minimal impact on targeting profitability — leaving limited room for improvement due to the scarcity of consequential customers.

mation errors, hindering the algorithm's ability to identify consequential customers. Ultimately, we recommend adopting a two-stage design with a larger proportion of customers sampled in the first stage, especially when the overall sample size is small.

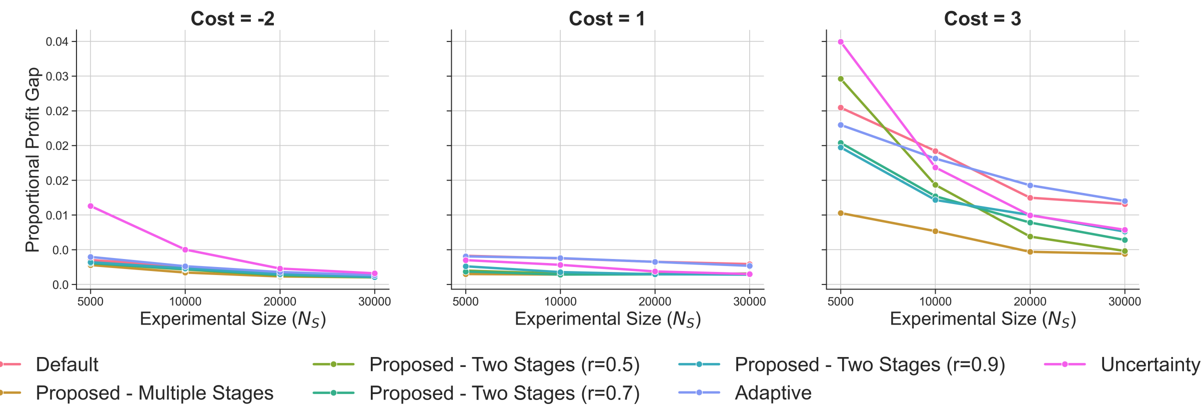


Figure W-9: Proportional Profit Gaps of Different Experimental Designs: Bimodal Distribution with Two Unequal Segments
We report the average value of the proportional profit gap across 100 replications. Each line corresponds to an experimental approach.

Figure W-10 further shows the proportional profit gaps of the targeting policies learned by different experimental designs under different sample allocation schemes. The results are qualitatively similar to the one with the baseline fixed batch size of 200. In particular, our approach consistently outperforms the three benchmarks when the number of customers around the decision threshold is limited (i.e., $c = -1$). Notably, the performance of our approach remains similar across both alternative allocation schemes (constant and decreasing batch sizes), indicating that the specific allocation strategy has minimal impact on the effectiveness of our approach.

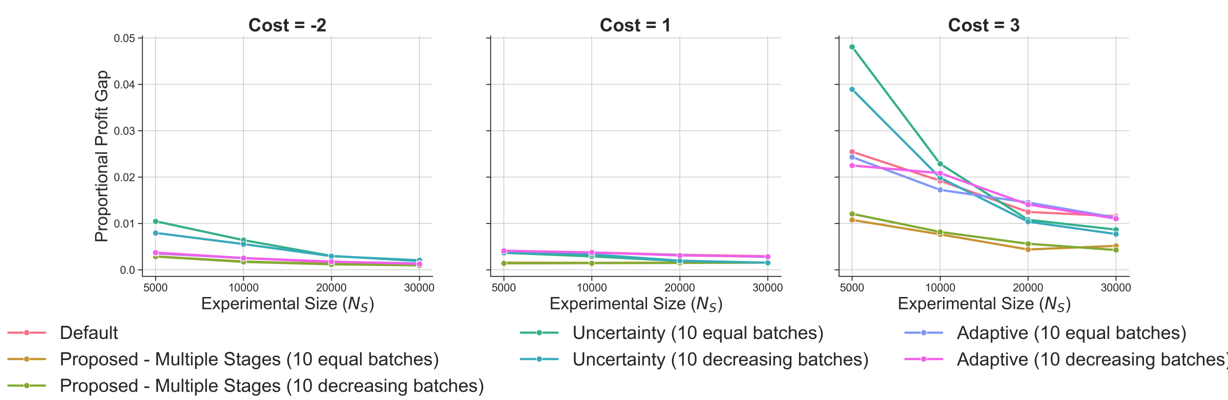


Figure W-10: Proportional Profit Gaps of Different Experimental Designs under Alternative Sample Allocation Schemes
We report the average value of the proportional profit gap across 100 replications. Each line corresponds to an experimental approach.

Figure W-11 illustrates the CATE distributions of sampled customers across varying intervention costs for different approaches. The findings highlight how our method effectively identifies and targets consequential customers with heightened intensity.

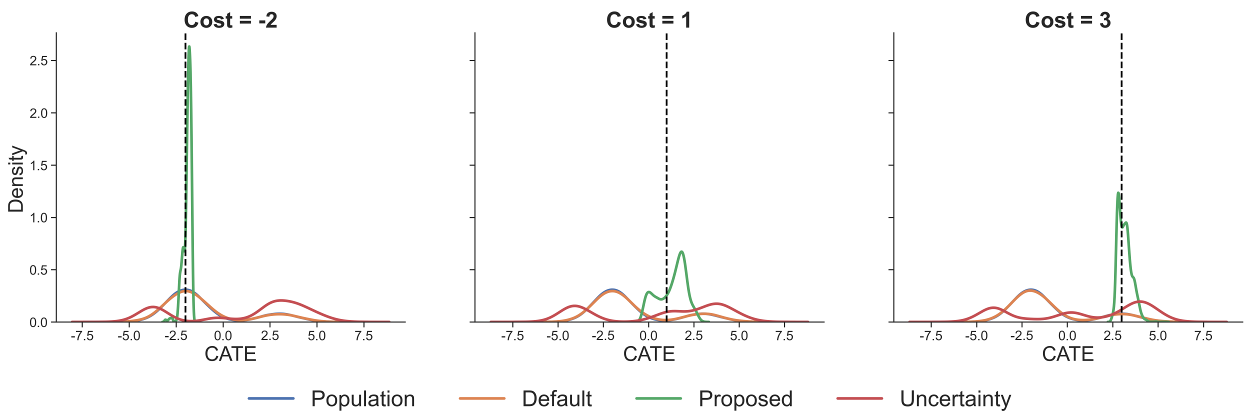


Figure W-11: CATE Distributions of Customers Sampled by Different Approaches: Bimodal Distribution with Two Unwqual Segments

Each line corresponds to the CATE distribution of the customers sampled by different approaches. The dashed line represents the intervention cost, which is also the decision threshold.

D. Further Details of Empirical Application

D.1. Summary Statistics and Randomization Check for Telecommunication Dormant Reactivation Campaign

Table W-2 presents the summary statistics of the pre-treatment covariates for the telecommunication campaign data. Additionally, we perform a weekly randomization check to verify the correct implementation of the randomization process.[33] The results indicate proper randomization, with no significant differences observed between the treatment and control groups across most variables and weeks.

D.2. Summary Statistics and Randomization Check for Starbucks Promotional Campaign

Table W-3 presents the summary statistics of the pre-treatment covariates for the Starbucks data. We also conduct randomization check to verify the correct implementation of the randomization process. The results suggests proper randomization, as there are no significant differences between the treatment and control groups in most of the variables.

D.3. Implementation Details

D.3.1. Default (A/B Test with Random Sampling)

For each experimental size N_S, we randomly sample N_S customers from the customer base $\mathcal{I}$. Since the treatment assignments in the original data are properly randomized, we use the treatment assignment from the original data as the final treatment assignment for each sampled customer in our experimentation. For CATE estimation, we construct a Causal Forest model (Wager and Athey 2018) implemented using the `econML` package in Python. This model consists of 300 trees with a maximum depth of 5 to prevent overfitting.

D.3.2. Uncertainty Sampling

For each experimental size N_S, we partition the full sample into $\frac{N_S}{500}$ batches, each containing 500 customers, to streamline the evaluation process. Additionally, we evaluate alternative sample allocation schemes (constant batch size and decreasing batch size over 10 batches) to assess their

[33] Since customers were randomized on a weekly basis, the randomization check is conducted at the weekly level.

Table W-2: Summary Statistics of Telecommunication Dormant Reactivation Campaign

Variable	Type	Mean	Std.	Median
cancellation usage (7 days)	Continuous	-0.953	11.043	0.0
cancellation usage (14 days)	Continuous	-1.971	22.152	0.0
cancellation usage (30 days)	Continuous	-4.562	47.362	0.0
dawli usage (7 days)	Continuous	0.952	8.049	0.0
dawli usage (14 days)	Continuous	2.224	13.241	0.0
dawli usage (30 days)	Continuous	10.004	31.863	0.0
recharge (7 days)	Continuous	0.449	5.830	0.0
recharge (14 days)	Continuous	2.017	13.950	0.0
recharge (30 days)	Continuous	16.778	51.037	0.0
rental usage (7 days)	Continuous	1.169	11.725	0.0
rental usage (14 days)	Continuous	2.433	23.419	0.0
rental usage (30 days)	Continuous	6.419	50.742	0.0
total usage (7 days)	Continuous	1.479	9.483	0.0
total usage (10 days)	Continuous	2.205	12.266	0.0
total usage (14 days)	Continuous	3.532	16.132	0.0
total usage (30 days)	Continuous	18.236	45.200	1.050
transfer fee usage (7 days)	Continuous	0.002	0.052	0.0
transfer fee usage (14 days)	Continuous	0.004	0.104	0.0
transfer fee usage (30 days)	Continuous	0.019	0.396	0.0
general usage (7 days)	Continuous	0.310	3.417	0.0
general usage (14 days)	Continuous	0.842	6.081	0.0
general usage (30 days)	Continuous	6.355	25.405	0.0

impact on performance. For customers within each batch b, we use the original treatment assignment in the data as their final treatment assignment.

For uncertainty estimation, as well as for the final CATE estimation, we utilize a Causal Forest model implemented using the `econML` package in Python. This model consists of 300 trees with maximum depths not exceeding 5. We estimate the uncertainty associates with the CATE estimates $\sigma(\hat{\tau}_{S^{b-1}}(x))$ by computing the standard deviation across the predictions generated by each tree.

Table W-3: Summary Statistics of Starbucks Promotional Campaign

Variable	Type	Mean	Std.	Median
V1 = 0	Discrete	0.1256	–	–
V1 = 1	Discrete	0.3757	–	–
V1 = 2	Discrete	0.3735 –	–	
V1 = 3	Discrete	0.1252	–	–
V2	Continuous	29.9779	5.0009	29.9796
V3	Continuous	0.0	1.0	-0.0395
V4 = 1	Discrete	0.3200 –	–	
V4 = 2	Discrete	0.6800 –	–	
V5 = 1	Discrete	0.1837	–	–
V5 = 2	Discrete	0.3693	–	–
V5 = 3	Discrete	0.3855	–	–
V5 = 4	Discrete	0.0615	–	–
V6 = 1	Discrete	0.2491	–	–
V6 = 2	Discrete	0.2490	–	–
V6 = 3	Discrete	0.2508	–	–
V6 = 4	Discrete	0.2510	–	–
V7 = 1	Discrete	0.2975	–	–
V7 = 2	Discrete	0.7025	–	–

D.3.3. Proposed Approach with Multiple Stages

For each experimental size N_S, we partition the full sample into $\frac{N_S}{500}$ batches, each containing 500 customers, to streamline the evaluation process. Additionally, we evaluate alternative sample allocation schemes (constant batch size and decreasing batch size over 10 batches) to assess their impact on performance. For customers within each batch b, we use the original treatment assignment in the data as their final treatment assignment.

For EPL estimation, as well as for the final CATE estimation, we utilize a Causal Forest model implemented using the `econML` package in Python. This model consists of 300 trees with maximum depths not exceeding 5.

D.3.4. Proposed Approach with Two Stages

For each experimental size N_S, we consider three different proportions of customers r to sample in the first stage ($r \in \{0.5, 0.7, 0.9\}$) and follow a two-stage sampling approach:

1. In the first stage, we randomly sample $r \cdot N_S$ customers from the customer base $\mathcal{I}$.
2. In the second stage, we sample the remaining $(1 - r) \cdot N_S$ customers who have the highest EPL estimated from the first stage.
3. For customers sampled in both stages, we use the original treatment assignment in the data as their final treatment assignment.

For EPL and final CATE estimation, we employ a Causal Forest model implemented using the `econML` package in Python. This model comprises 300 trees with maximum depths not exceeding 5.

D.4. Additional Results for Alternative Sample Allocation Schemes

In this appendix, we present additional results examining alternative sample allocation schemes for the two empirical applications. Specifically, for the multi-stage design, we evaluate two sample allocation approaches: constant batch size and decreasing batch size over 10 batches. In addition, we explore three different configurations of the two-stage design by varying the proportions of customers r to sample in the first stage ($r \in \{0.5, 0.7, 0.9\}$) and compare their performance with the multi-stage design.

D.4.1. Alternative Sample Allocation Schemes for Multi-Stage Design

Figure W-12 and Figure W-13 present the results of targeting policies learned under different sample allocation schemes for the two empirical applications, respectively. The results are qualitatively similar to those obtained with the baseline fixed batch size of 500. In particular, our approach consistently outperforms the default method when the intervention creates a risk of cannibalizing profits that would have been earned without intervention (the telecommunication application and the 50% discount scenario for Starbucks). For uncertainty sampling, our approach performs comparably in the telecommunication case due to high overlap in selected customers, but outperforms it in 50% discount scenario for Starbucks application. Moreover, both allocation schemes yield similar performance across scenarios, providing further evidence that the specific allocation strategy has minimal impact on the effectiveness of our approach.

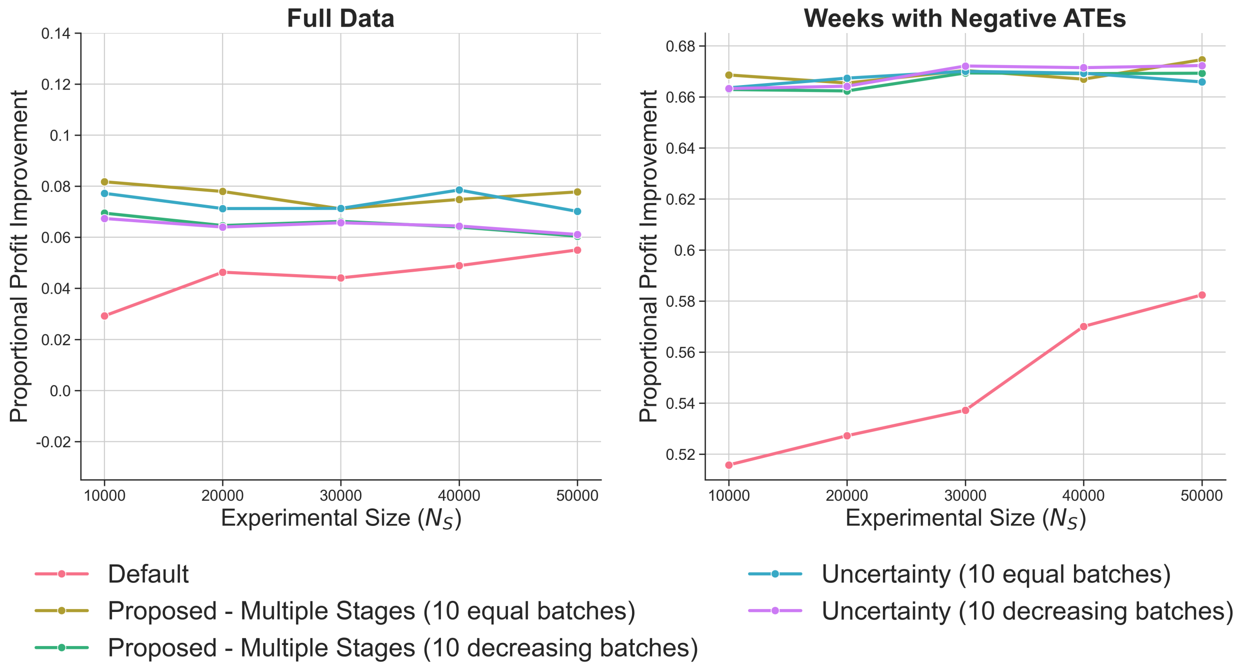


Figure W-12: Performance of Targeting Policies Learned from Different Multi-Stage Designs (Telecommunication)
We report the average value of the proportional profit improvement relative to the uniform policy across 100 replications. Each line corresponds to an experimental approach.

D.4.2. Alternative Configurations for Two-Stage Design

Figure W-14 and Figure W-15 present the results of targeting policies learned by a multi-stage design and three two-stage design of our approach under different configurations ($r \in \{0.5, 0.7, 0.9\}$). In general, the proportion of customers sampled in the first stage does not substantially impact the overall performance of the two-stage design.

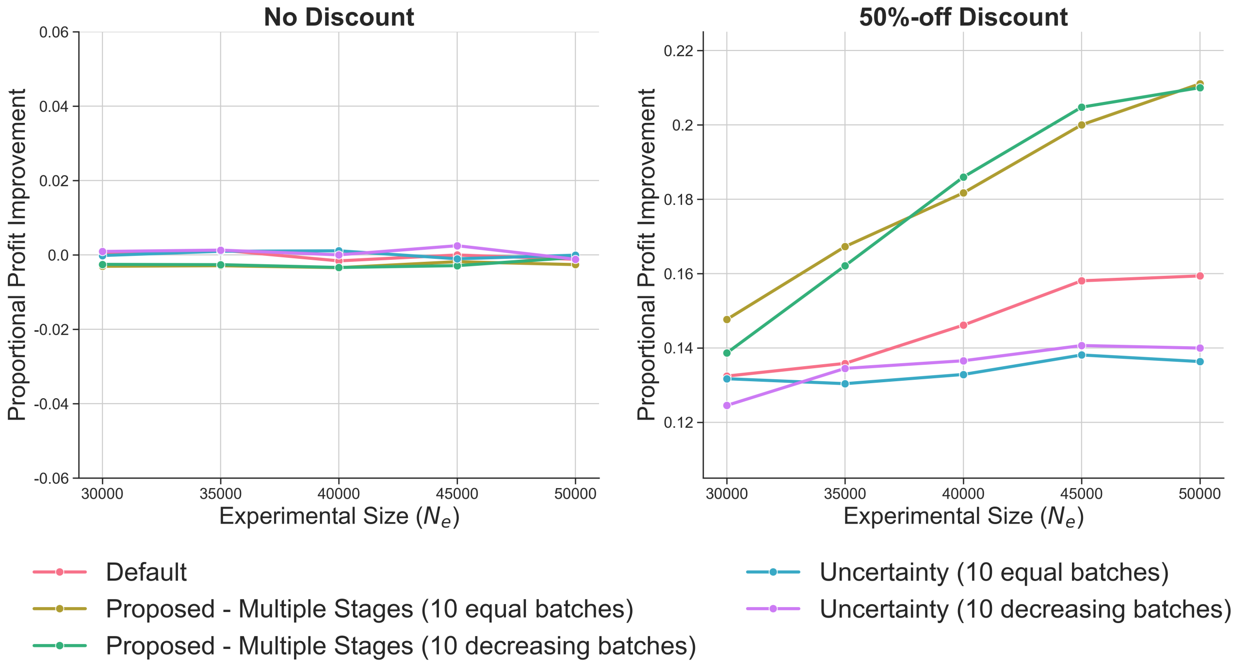


Figure W-13: Performance of Targeting Policies Learned from Different Multi-Stage Designs (Starbucks)

We report the average value of the proportional profit improvement relative to the uniform policy across 100 replications. Each line corresponds to an experimental approach.

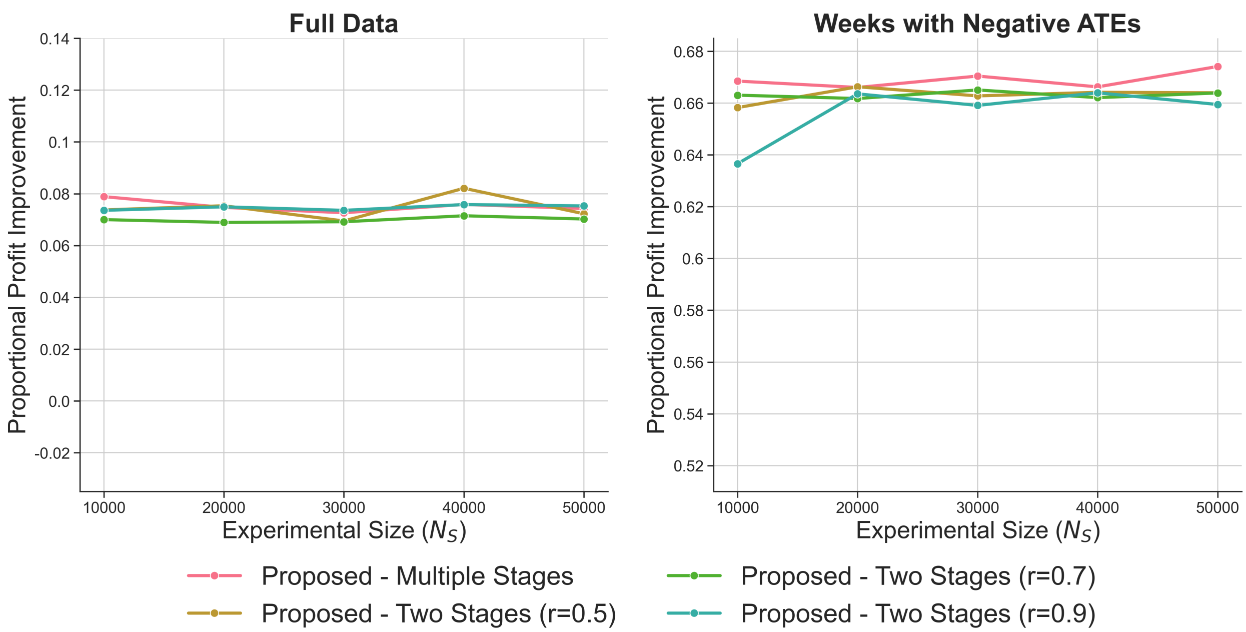


Figure W-14: Performance of Targeting Policies Learned from Different Two-Stage Designs (Telecommunication)

We report the average value of the proportional profit improvement relative to the uniform policy across 100 replications. Each line corresponds to an experimental approach.

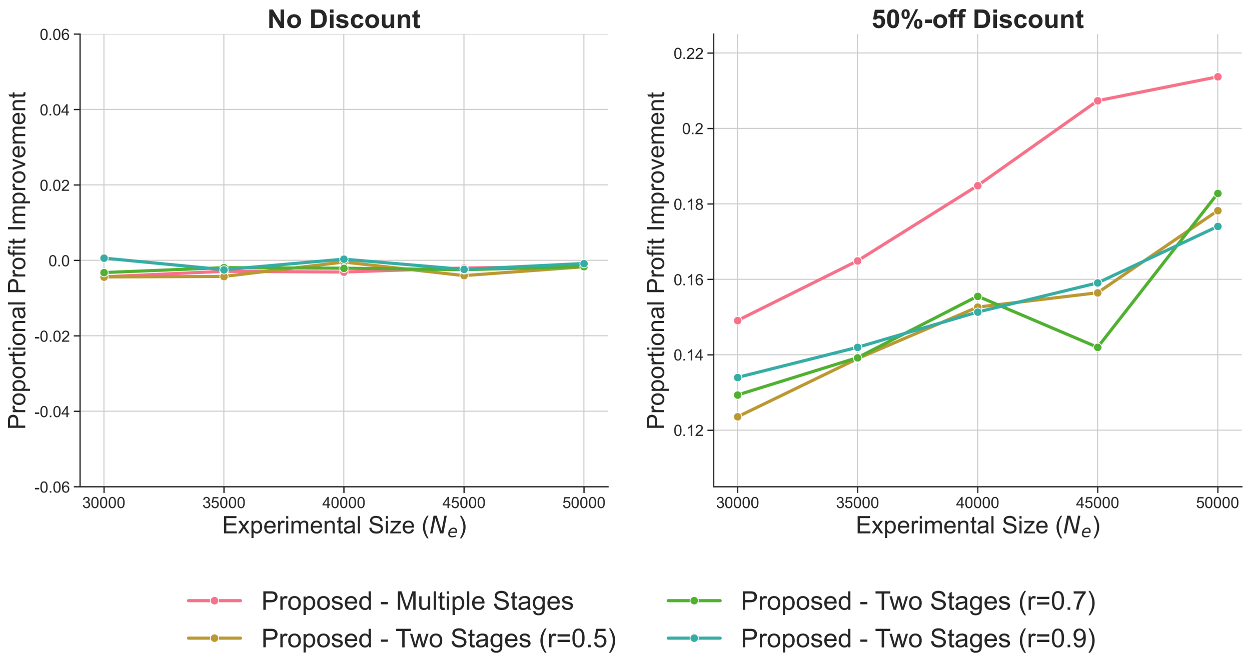


Figure W-15: Performance of Targeting Policies Learned from Different Two-Stage Designs (Starbucks)

We report the average value of the proportional profit improvement relative to the uniform policy across 100 replications. Each line corresponds to an experimental approach.

References

Aramayo, N., Schiappacasse, M., and Goic, M. (2022). A Multiarmed Bandit Approach for House Ads Recommendations. *Marketing Science*, page mksc.2022.1378.

Athey, S., Byambadalai, U., Hadad, V., Krishnamurthy, S. K., Leung, W., and Williams, J. J. (2022). Contextual Bandits in a Survey Experiment on Charitable Giving: Within-Experiment Outcomes versus Policy Learning. *arXiv preprint arXiv:2211.12004*. arXiv:2211.12004 [cs, econ, stat].

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2).

Athey, S. and Wager, S. (2021). Policy Learning With Observational Data. *Econometrica*, 89(1):133–161.

Bubeck, S., Munos, R., and Stoltz, G. (2010). Pure Exploration for Multi-Armed Bandit Problems. *arXiv preprint arXiv:0802.2655*. arXiv:0802.2655 [cs, math, stat].

Cardoso, T. N., Silva, R. M., Canuto, S., Moro, M. M., and Gonçalves, M. A. (2017). Ranked batch-mode active learning. *Information Sciences*, 379:313–337.

Caria, S., Gordon, G., Kasy, M., Quinn, S., Shami, S., and Teytelboym, A. (2020). An Adaptive Targeted Field Experiment: Job Search Assistance for Refugees in Jordan. *Available at SSRN 3689456*.

Carranza, A. G., Krishnamurthy, S. K., and Athey, S. (2023). Flexible and Efficient Contextual Bandits with Heterogeneous Treatment Effect Oracles. *arXiv preprint arXiv:2203.16668*. arXiv:2203.16668 [cs, math, stat].

Chick, S. E. and Frazier, P. (2012). Sequential Sampling with Economics of Selection Procedures. *Management Science*, 58(3):550–569.

Feit, E. M. and Berman, R. (2019). Test & Roll: Profit-Maximizing A/B Tests. *Marketing Science*, 38(6):1038–1058.

Fontaine, X., Perrault, P., Valko, M., and Perchet, V. (2020). Online A-Optimal Design and Active Linear Regression. arXiv:1906.08509 [stat].

Fu, Y., Zhu, X., and Li, B. (2013). A survey on instance selection for active learning. *Knowledge and Information Systems*, 35(2):249–283.

Grover, A., Markov, T., Attia, P., Jin, N., Perkins, N., Cheong, B., Chen, M., Yang, Z., Harris, S., Chueh, W., and Ermon, S. (2018). Best arm identification in multi-armed bandits with delayed feedback. *arXiv preprint arXiv:1803.10937*. arXiv:1803.10937 [cs, stat].

Hauser, J. R., Urban, G. L., Liberali, G., and Braun, M. (2009). Website Morphing. *Marketing Science*, 28(2):202–223.

Hu, Y., Zhu, H., Brunskill, E., and Wager, S. (2024). Minimax-Regret Sample Selection in Randomized Experiments. *arXiv preprint arXiv:2403.01386*. arXiv:2403.01386 [econ, stat].

Jedra, Y. and Proutiere, A. (2020). Optimal Best-arm Identification in Linear Bandits. *arXiv preprint arXiv:2006.16073*. arXiv:2006.16073 [cs, stat].

Jesson, A., Tigas, P., van Amersfoort, J., Kirsch, A., Shalit, U., and Gal, Y. (2022). Causal-BALD: Deep Bayesian Active Learning of Outcomes to Infer Treatment-Effects from Observational Data. *arXiv preprint arXiv:2111.02275*. arXiv:2111.02275 [cs, stat].

Kasy, M. and Sautmann, A. (2021). Adaptive Treatment Assignment in Experiments for Policy Choice. *Econometrica*, 89(1):113–132.

Kato, M., Okumura, K., Ishihara, T., and Kitagawa, T. (2024). Adaptive Experimental Design for Policy Learning. *arXiv preprint arXiv:2401.03756*. arXiv:2401.03756 [cs, econ, stat].

Kiefer, J. and Wolfowitz, J. (1959). Optimum Designs in Regression Problems. *The Annals of Mathematical Statistics*, 30(2):271–294.

Li, L., Chu, W., Langford, J., and Schapire, R. E. (2010). A Contextual-Bandit Approach to Personalized News Article Recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670. arXiv:1003.0146 [cs].

Liu, M., Grigas, P., Liu, H., and Shen, Z.-J. M. (2023). Active Learning in the Predict-then-Optimize Framework: A Margin-Based Approach. *arXiv preprint arXiv:2305.06584*. arXiv:2305.06584 [cs, math, stat].

Misra, K., Schwartz, E. M., and Abernethy, J. (2019). Dynamic Online Pricing with Incomplete Information Using Multiarmed Bandit Experiments. *Marketing Science*, 38(2):226–252.

Nemhauser, G. L., Wolsey, L. A., and Fisher, M. L. (1978). An analysis of approximations for maximizing submodular set functions—I. *Mathematical Programming*, 14(1):265–294.

Puha, Z., Kaptein, M., and Lemmens, A. (2020). Batch Mode Active Learning for Individual Treatment Effect Estimation. In *2020 International Conference on Data Mining Workshops (ICDMW)*, pages 859–866, Sorrento, Italy. IEEE.

Robinson, P. M. (1988). Root-N-Consistent Semiparametric Regression. *Econometrica*, 56(4):931.

Schwartz, E. M., Bradlow, E. T., and Fader, P. S. (2017). Customer Acquisition via Display Advertising Using Multi-Armed Bandit Experiments. *Marketing Science*, 36(4):500–522.

Simester, D., Timoshenko, A., and Zoumpoulis, S. (2022). A Sample Size Calculation for Training and Certifying Targeting Policies. *Available at SSRN 4228297*.

Sundin, I., Schulam, P., Siivola, E., Vehtari, A., Saria, S., and Kaski, S. (2019). Active Learning for Decision-Making from Imbalanced Observational Data. *arXiv preprint arXiv:1904.05268*. arXiv:1904.05268 [cs, stat].

Wager, S. and Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, 113(523):1228–1242.

Waisman, C., Nair, H. S., and Carrion, C. (2024). Online Causal Inference for Advertising in Real-Time Bidding Auctions. *Marketing Science*, page mksc.2022.0406.

Wang, Z. and Ye, J. (2015). Querying Discriminative and Representative Samples for Batch Mode Active Learning. *ACM Transactions on Knowledge Discovery from Data*, 9(3):1–23.