Policy-Aware Sampling: Prioritizing Consequential Customers 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

March 23, 2026

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 suboptimal targeting policies. We propose a *policy-aware sampling* strategy — that directly incorporates the firm's profit-maximizing objective into the selection of which customers to sample in the experiment. Specifically, we introduce *expected profit loss (EPL)*, a criterion that quantifies customers' learning value by measuring the financial impact of treatment effect estimation errors on targeting profitability. We show that allocating experimental resources based on EPL achieves near-optimality with theoretical guarantees, and develop a sequential sampling algorithm that prioritizes these consequential 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 benchmark methods, and achieves comparable outcomes with up to 97% fewer experimental samples, highlighting substantial gains in both targeting effectiveness and data efficiency. Our results underscore the importance of aligning experimental design and accuracy considerations with business objectives.

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, 2025 Conference on Artificial Intelligence, Machine Learning, and Business Analytics, and the 42^{nd} University of Houston Doctoral Symposium for valuable feedbacks. The authors received no financial support for the research, authorship, or publication of this article and declare no conflicts of interest. Generative AI (Claude) was used solely to correct grammar and improve writing clarity; all ideas, analyses, and content are entirely the authors' own.

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 risk 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 key 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.

As a result, firms may inefficiently allocate experimental resources, undersampling customers whose data are most consequential for determining optimal targeting policies. 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 and adapt them to the specific structure of profit-maximizing targeting problems commonly studied in marketing. While prior work in this stream of literature focuses on reducing predictive uncertainty or cumulative regret, firms engaged in targeting face a distinct learning problem: identifying customers whose estimation errors can flip targeting decisions and lead to economically meaningful losses. To tackle this issue, we introduce a profit-based sampling criterion — *expected profit loss (EPL)* — which quantifies customers' learning value based on how estimation errors can distort targeting decisions and lead to economically meaningful losses. EPL prioritizes customers for whom uncertainty in treatment effect estimation is most likely to distort the firm's targeting rule and generate profit losses. Unlike standard active learning methods that prioritize observations that improve global predictive accuracy or reduce posterior variance, our EPL-based sampling approach focuses learning resources on the consequential customers to improve targeting profitability.

We theoretically show that, within a class of threshold-based targeting policies, the EPL-based sampling approach achieves near-optimality with a provable performance guarantee, and develop an estimation strategy leveraging the theoretical properties of Causal Forest estimators (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 our approach in two real-world applications — a telecommunications reactivation campaign and a Starbucks promotional offer — as well as extensive simulation studies. We empirically demonstrate that the proposed policy-aware sampling approach yields more profitable targeting policies than conventional sampling designs, including standard test-then-learn, uncertainty sampling, and a state-of-the-art adaptive design (e.g., Kato et al. 2024). Across these empirical settings, the proposed approach improves profitability by 5% to 10% relative to standard methods, and achieves comparable outcomes with up to 97% fewer experimental samples. These results highlight both the economic relevance and the sample efficiency of our method, especially relevant in settings where experimental budgets are limited.

The proposed method's advantage over alternative sampling strategies arises particularly in settings where customers with less predictable behavior (e.g., infrequent buyers, new customers) are also more responsive to interventions, as demonstrated in our telecommunication application, and when 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). These scenarios highlight the practical relevance of our approach in real-world marketing contexts where such conditions frequently arise.

This research makes three key contributions. First, we formalize expected profit loss (EPL), a profit-based value-of-information sampling criterion tailored to targeting decisions, and show that it achieves near-optimality by prioritizing sampling customers with the greatest economic impact for personalized targeting. Second, we show that experimentally oversampling economically consequential customers — rather than uniformly reducing estimation error — can substantially improve targeting profitability without increasing sample size. Third, we provide a practical experimental design that preserves the simplicity of test-then-learn workflows while incorporating decision-aware sampling, lowering barriers to adoption for firms.

2. Problem Formulation

Consider a firm planning a reactivation campaign on its existing customers. In particular, the firm is deciding whether to send a $2 coupon to each customer.[1] 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 experimental data. As a result, targeting decisions are based on predictions, which inevitably include prediction error.

To illustrate the implications of prediction error, consider three customers (A, B, and C) in Figure 1, who should be targeted if their true CATEs exceed the $2 threshold. Figure 1a shows a (homoskedastic) case where all three customers have identical prediction errors. For customer C, whose true CATE is far above $2, prediction errors are relatively inconsequential as the firm will likely make the correct targeting decision despite the error. 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, making customer B the most consequential.

In contrast, Figure 1b shows a heteroskedastic case where larger treatment effects tend to come with greater uncertainty. This arises naturally in practice: customers who are most responsive to

[1] We focus on scenarios where the firm has access to a fixed customer base with observed characteristics and can strategically select which customers to experiment on, excluding cases with streaming arrivals. These scenarios are common in a wide range of 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).

interventions are often those whose baseline behavior is least stable (e.g., infrequent buyers, new customers), making them simultaneously harder to predict. Here, customer C has the largest prediction error, which causes mistargeting with the largest profit loss, making customer C most consequential.

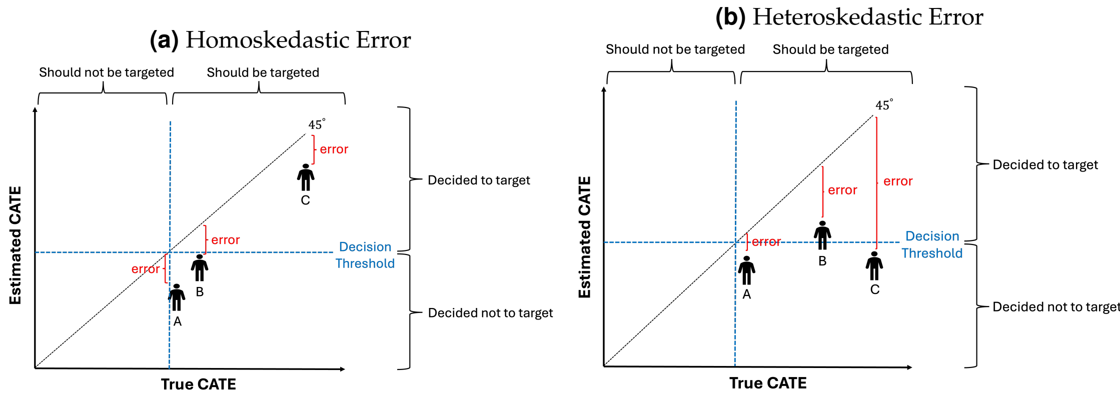


Figure 1: Differential Impact of Prediction Errors on Targeting Profitability
Prediction errors matter most when they cause incorrect targeting decisions with large profit implications: customer B in (a), and customer C in (b). Note that the argument holds irrespective of the direction of error, as only whether errors cross the decision threshold matters. For illustration, we show cases where estimated CATE underestimates true CATE, but the same principles apply for overestimation.

Figure 1 highlights a key insight: when firms rely on CATE estimates to deploy targeted policies, not all customers contribute equally to the quality of decisions. In particular, prediction errors are more consequential for some customers than for others. This insight motivates our experimental design approach: sampling effort should be concentrated on these consequential customers rather than allocated uniformly or based solely on predictive uncertainty. This intuition directly underlies the policy-aware sampling strategy introduced in Section 1

2.1. Formal Setup

To formalize this problem, we consider a setting where the firm has a pre-determined customer base $\mathcal{I}$ with each customer $i \in \mathcal{I}$ 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 $\mathbf{X}_i$ and $c(\cdot)$ is the intervention cost.[2]

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^*

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

only when estimation is sufficiently accurate. Second, since the policy π is implemented via a threshold rule that treats 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. This observation raises a natural design question: how should firms choose which customers to experiment on when learning resources are limited?

Policy-Aware Sampling We formalize that objective as follows: given a fixed budget for sampling N_S customers into the experiment (i.e., $|S| = N_S$) — reflecting common budget or operational constraints in marketing experiments — 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 the entire customer base ($\mathcal{I}$):[3]

$$S^* = \max_{S \in \mathcal{I}, |S| = N_S} f(S) \tag{4}$$

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

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 $\binom{|\mathcal{I}|}{N_S}$ from a large population quickly becomes computationally infeasible. 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 sampling

[3]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).

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 the proposed 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.

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; Hitsch et al. 2024; Huang and Ascarza 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 (e.g. Zhao et al. 2012; Athey and Wager 2021). 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. Within the marketing targeting literature, prior work has largely treated experimental design as separate from downstream business objectives, leaving a 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 organize into two main streams. The first stream focuses on non-adaptive experimental design, either with the objective of minimizing parameter uncertainty for global quantities like average treatment effects using parametric models such as classical A- and D-optimal designs (e.g. Kiefer and Wolfowitz 1959; Fontaine et al. 2020), or determining the optimal experimental size to achieve a specific objective (Blattberg 1979; Ginter et al. 1981; Feit and Berman 2019; Simester et al. 2022). The works most closely aligned with ours include Toubia and Hauser (2007), who pioneered managerially efficient designs that accounts for downstream decision making, and Hu et al. (2024), who focuses on optimizing sampling procedures for better decision quality. Our work differs in two key ways: First, we address cases where modeling heterogeneity in treatment effect estimation is essential, unlike Toubia and Hauser (2007) who focus on uniform policies that treat all customers identically. Second, our method does not rely on any pre-segmentation of customers as in Hu et al. (2024), making it applicable in various business contexts where companies are uncertain about the true and relevant customer segmentation for the business problem at hand.

The second stream includes adaptive experimental methods drawing from bandits (e.g., Hauser et al. 2009; Chick and Frazier 2012; Ryzhov et al. 2012; Urban et al. 2014; Misra et al. 2019; Waisman et al. 2024) and best arm identification (e.g., Bubeck et al. 2010; Grover et al. 2018; Jedra and Proutiere 2020; Kasy and Sautmann 2021; Carranza et al. 2023; Kato et al. 2024). Unlike bandit algorithms that balance exploration with exploitation to minimize cumulative regret, which requires limiting sampling of suboptimal arms, we focus on minimizing a simple regret (i.e., the quality of the final learned policy), which requires sufficient sampling across the covariate space to learn the CATE function accurately. As Bubeck et al. (2010) demonstrate, these objectives are fundamentally different, as strategies that minimize cumulative regret can lead to larger simple regret. Our method, similar to best arm identification, focuses purely on optimizing the exploration phase without concern for rewards accumulated during experimentation.

While our approach shares similarities with best arm identification (BAI) methods, it shifts the focus from assigning treatment conditions to selecting *whom* to sample in order to improve downstream decision quality. This customer selection dimension provides an additional degree of freedom that is absent in BAI methods, and strategically exploiting it can substantially improve the quality of learned targeting policies beyond what can be achieved through treatment assignment optimization alone. Our simulation results demonstrate that this customer selection approach can yield substantial gains relative to state-of-the-art BAI methods, highlighting the practical value of this additional design dimension.

Active learning. Third, our research is closely related to the active learning literature, which develops acquisition strategies for selectively choosing which observations to sample or label in order 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; Filstroff et al. 2021). Related ideas of resource allocation based on proximity to decision thresholds and measurement uncertainty also appear in other domains, such as educational testing (e.g., Bradlow and Wainer 1998), where raters are allocated based on distance from cutscores, potential loss, and error in measurement.

Our approach differs from existing uncertainty-based acquisition functions in an important way: rather than prioritizing cases based solely on the probability of misclassification (as in Sundin et al. 2019; Filstroff et al. 2021), our expected profit loss (EPL) criterion recognizes that firms care about total profitability, not merely decision correctness. By weighting prediction uncertainty by treatment effect magnitude, EPL prioritizes sampling where errors are economically costly. This distinction proves especially valuable in business settings where the customers hardest to predict — such as new or infrequent buyers — are also those most likely to respond strongly to marketing interventions, as demonstrated in our empirical applications.

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 in supervised learning settings (e.g., Elmachtoub and Grigas 2022; Chung et al. 2022), recent work has begun to address data collection strategies within this framework (e.g., Liu et al. 2023). Our work extends this line of research to the more challenging context of counterfactual reasoning, where the outcome of interest — CATEs — are unobserved rather than directly observable as in supervised settings. In doing so, we offer a marketing-focused perspective on aligning experimental design with business objectives.

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), a profit-based value-of-information measure that quantifies the learning value of including a customer in the testing phase by measuring 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 the theoretical properties of Causal Forest estimators (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 consequential customers of the population.

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

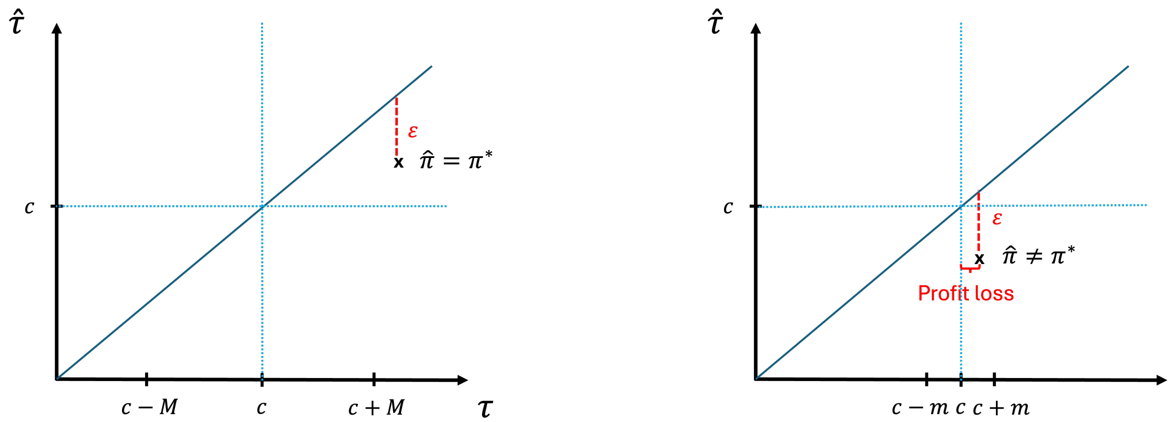
4.1. Near-Optimal Sampling Strategy: The Expected Profit Loss (EPL) Sampling Approach

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.



(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

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

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

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, and let N_S denote the desired total sample size (with $N_S > |\mathcal{D}|$). Define the objective function as the expected profit from deploying the learned targeting policy $\hat{\pi}_S$ on future customers:*

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

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

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 that maximizes $f(\mathcal{D} \cup S_k)$:

$$S_k^* = \arg \max_{S_k \subseteq \mathcal{I},\, |S_k| = k} f(\mathcal{D} \cup S_k). \tag{6}$$

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 in threshold-based targeting settings under a fixed experimental budget, offering theoretical performance guarantees of our proposed approach. Importantly, this guarantee applies within the class of fixed-budget sampling policies for threshold-based targeting; extensions to adaptive treatment assignment or more general policy-learning settings are beyond the scope of this analysis.

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. In practice, as we discussed later in Section 4.3, our simulation results indicate that initial samples on the order of a few hundred customers are sufficient to stabilize EPL estimates and enable effective prioritization.

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. 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 the asymptotic normality of Causal Forest estimators (Wager and Athey 2018; Athey et al. 2019b) to avoid restrictive parametric assumptions on CATEs. Specifically, for an asymptotically normal estimator $\hat{\tau}_{\mathcal{D}}(\mathbf{X}_i)$ estimated from a dataset $\mathcal{D}$, the confidence distribution of the parameter $\tau(\mathbf{X}_i)$ can be expressed as:

$$h_{\mathcal{D}}(\tau(\mathbf{X}_i)) = \frac{1}{\hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)}\phi\left(\frac{\tau(\mathbf{X}_i) - \hat{\tau}_{\mathcal{D}}(\mathbf{X}_i)}{\hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)}\right)$$

where ϕ is the density function of the standard normal distribution and $\hat{\sigma}(\mathbf{X}_i)$ is the consistent estimator of the standard error $\sigma(\mathbf{X}_i)$. This confidence distribution provides a distributional approximation to $\tau(\mathbf{X}_i)$ that provides sufficient summary of the information currently available about $\tau(\mathbf{X}_i)$, similar to the style of a Bayesian posterior (Xie and Singh 2013), which we use to integrate profit loss over estimation uncertainty.

We then define the EPL for customer i as the expected profit loss arising from discrepancies between the targeting decision based on our current point estimate $\hat{\tau}_{\mathcal{D}}(\mathbf{X}_i)$ and the targeting decisions that would be made under alternative realizations of $\tau(\mathbf{X}_i)$ drawn from the confidence distribution. Formally,

$$\begin{aligned}\hat{\ell}_{\mathcal{D}}(\mathbf{X}_i) &= \int \underbrace{0 \cdot \mathbf{1}\{\hat{\pi}_{\mathcal{D}}(\mathbf{X}_i) = \pi(\mathbf{X}_i)\}}_{\text{profit loss with no deviation}} + \underbrace{|\tau(\mathbf{X}_i) - c(\mathbf{X}_i)|\mathbf{1}\{\hat{\pi}_{\mathcal{D}}(\mathbf{X}_i) \neq \pi(\mathbf{X}_i)\}}_{\text{profit loss with deviation}} h_{\mathcal{D}}(\tau(\mathbf{X}_i))d\tau \\ &= \int \underbrace{|\tau(\mathbf{X}_i) - c(\mathbf{X}_i)|\mathbf{1}\{\hat{\pi}_{\mathcal{D}}(\mathbf{X}_i) \neq \pi(\mathbf{X}_i)\}}_{\text{profit loss}} h_{\mathcal{D}}(\tau(\mathbf{X}_i))d\tau \qquad (7) \\ &= \hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)\left(\phi(r(\mathbf{X}_i)) - r(\mathbf{X}_i)\Phi(-r(\mathbf{X}_i))\right)\end{aligned}$$

where

$$r(\mathbf{X}_i) = \frac{|\hat{\tau}_{\mathcal{D}}(\mathbf{X}_i) - c(\mathbf{X}_i)|}{\hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)}$$

and Φ is the cumulative distribution function of the standard normal (see detailed derivation in Web Appendix C).

To implement this estimation in practice, we first use the previously collected dataset $\mathcal{D}$ to train a Causal Forest model, obtaining both $\hat{\tau}_{\mathcal{D}}(\mathbf{X}_i)$ (the CATE estimate) and $\hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)$ (the standard error estimate) for each customer. The standard error is estimated using the bootstrap-of-little-bag approach (Athey et al. 2019b), which provides theoretical guarantees for consistency. Since Causal Forest provides asymptotically normal estimates of the true CATE $\tau(\mathbf{X}_i)$ with these consistent standard errors (Wager and Athey 2018; Athey et al. 2019b), we can directly compute the EPL by substituting these estimates into Equation (7). The estimation procedure is outlined in Algorithm 1.

Algorithm 1 Estimating Expected Profit Loss

Input: Experimental data $\mathcal{D}$; Cost threshold $c(\mathbf{X}_i)$
Output: $\hat{\ell}_{\mathcal{D}}(\mathbf{X}_i)$
Train a Causal Forest model on $\mathcal{D}$
Obtain CATE estimates $\hat{\tau}_{\mathcal{D}}(\mathbf{X}_i)$ for each customer i
Obtain standard error estimates $\hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)$ using the bootstrap-of-little-bag approach
Compute the normalized distance to threshold: $r(\mathbf{X}_i) = \frac{|\hat{\tau}_{\mathcal{D}}(\mathbf{X}_i) - c(\mathbf{X}_i)|}{\hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)}$
Calculate the EPL estimate using the closed-form solution:

$$\hat{\ell}_{\mathcal{D}}(\mathbf{X}_i) = \hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)\left(\phi(r(\mathbf{X}_i)) - r(\mathbf{X}_i)\Phi(-r(\mathbf{X}_i))\right)$$

where ϕ is the standard normal density function and Φ is the standard normal cumulative distribution function
Return: EPL estimate $\hat{\ell}_{\mathcal{D}}(\mathbf{X}_i)$

4.2.2. A Sequential Experimental Design with Expected Profit Loss Sampling

Having developed an estimation procedure for EPL, we next describe how these scores can be operationalized to guide data collection over the course of the experiment. An important consideration is that EPL itself depends on estimated treatment effects, which in turn require experimental data, creating a natural feedback loop between learning and customer selection. We therefore consider a sequential experimental design in which customer selection is updated as new data become available. In its idealized form, such a design begins with an initial random sample to obtain preliminary CATE estimates, then proceeds iteratively: (1) select the customer with the highest estimated EPL based on current CATE estimates, (2) observe the outcome for this customer, (3) update the CATE model with this new observation, and (4) repeat the selection process with refined EPL estimates. As more data is collected, the accuracy of EPL estimates improves, allowing the procedure to progressively focus on customers whose mistargeting would lead to substantial profit loss.

This fully sequential approach requires a sufficiently large initial sample to ensure that for any focal customer i with feature vector $\mathbf{X}_i$, there are enough neighbors in the feature space similar to $\mathbf{X}_i$. Additionally, this initial data is necessary to compute the first set of EPL scores that enables subsequent strategic sampling. To satisfy these requirements, we begin by collecting a random sample of customers, forming the initial dataset $\mathcal{D}$.[4]

[4]We assume firms lack prior knowledge about treatment effect heterogeneity in the focal customer population, necessitating random sampling for initialization. However, firms with relevant prior information (e.g., from previous campaigns) can integrate this knowledge using transfer learning techniques (e.g., Huang et al. 2024; Ibragimov et al. 2025). Such methods allow firms to initialize EPL estimates based on historical treatment effect estimates and their uncertainties, then apply our adaptive sampling procedure to refine targeting for the focal context.

A key practical constraint arises when there are delays between interventions and outcomes, as is common in many business settings such as promotional activities. In such cases, a fully sequential approach (selecting one customer at a time) would make experiments impractical. To address this constraint, we propose a multi-stage sequential approach in batches when outcome delays exist. This approach allows firms to run experiments within realistic time frames while maintaining most of the statistical benefits of EPL. Regardless of whether the fully sequential or batch approach is used, as more data is collected, the accuracy of EPL estimates improves, progressively refining the prioritization of consequential customers.

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 $\sum_{b=1}^{B} n_b = N_S$. 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 $W_i \in \{0, 1\}$.[5] After observing their outcomes, we use this data S_1 to estimate EPLs for the remaining customers.[6] 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 $S = \cup_{b=1}^{B} S_b$.[7]

Crucially, our design differs from commonly used bandit or BAI 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:

[5] The initial sample can also be collected by stratified sampling, as long as it ensures sufficient coverage of the covariate space.

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

[7] We also explore whether combining our EPL-based customer selection with strategic treatment assignment, which allocates sampled customers to treatment arms based on potential outcome uncertainty, can further enhance targeting performance (see Web Appendix D.4). We find negligible additional gains from this hybrid approach, confirming that strategic customer selection is the primary driver of our method's performance advantages, while the marginal benefit of adaptive treatment assignment is minimal once consequential customers are appropriately sampled.

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 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,*

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

Assumption 2. *(Stability) The distribution of potential outcomes for any customer i is invariant over 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 $\tau(\mathbf{X}_i)$.[8] 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; (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. Throughout the procedure, each customer can be selected at most once, which is necessary given our one-shot treatment setting, in which each customer is exposed to the intervention at most once, as repeated selection would imply carryover effects that are inconsistent with this objective and would bias treatment effect estimation.

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

[8]Note that under this design, the final 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.

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, especially when the desired outcomes (e.g., purchases, revenue) take days or weeks to manifest.[9] 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). We recommend setting B to 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 large number of batches may be practical. In contrast, when outcomes require longer observation windows (e.g., revenue after a promotion), a two-stage design with EPL sampling only occurring in the second stage may be more suitable. To understand the trade-off between simplicity and profit performance in such settings, we investigate the two-stage design in detail in Section 5 and 6. To add flexibility, firms can adopt an *early stopping rule* — terminating the experiment once EPL estimates stabilize or fall below a pre-specified threshold (see Web Appendix F.1 for details).

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. Nevertheless, our simulation (Web Appendix D.3.1) and empirical results (Web Appendix E.5.1) suggest that both constant and decreasing allocations perform similarly in practice.

A related practical question concerns how large the initial sample should be to satisfy the theoretical requirement and provide stable initialization. Our results demonstrate that the multi-stage sequential approach is robust to initial sample size choices. Even modest initial samples

[9] 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.

(e.g., 300 customers) suffice for effective sampling procedure as the iterative refinement across subsequent batches compensates for initial uncertainty. The decreasing batch configuration shows only marginal advantages at very small total experimental sizes.

On the other hand, when the firm adopts a simplified two-stage design, the selection of initial batch size becomes critical as it 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 Web Appendices D.3.2 and E.5.2.

Deterministic vs. stochastic sampling. A natural concern is whether the deterministic selection of highest-EPL customers might cause the algorithm to become trapped in suboptimal regions of the covariate space if initial exploration is insufficient. Our method inherently balances strategic sampling and coverage through its uncertainty-aware design: if a region is underrepresented in early samples, CATE estimates for customers in that region will have high standard errors, which directly increases their EPL scores (recall $\hat{\ell}_D(X_i) = \hat{\sigma}_D(X_i)(\phi(r) - r\Phi(-r))$ from Equation 7). Thus, customers from underexplored regions naturally receive priority in subsequent batches. To empirically validate this property, in Web Appendices D.5 and E.6 we compared our deterministic approach with a stochastic variant that selects customers with probability proportional to their EPL scores. We find negligible performance differences, indicating that our deterministic method provides sufficient exploration through uncertainty quantification alone.

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. In addition to the proposed (multi-stage) approach, we evaluate a simplified (two-stage) design, as discussed in Section 4.3. We find that this simplified design maintains strong performance while offering a more practical alternative to fully adaptive implementation. Second, we leverage the advantages of synthetic data where true CATE is known to evaluate the ability of our approach to accurately identify consequential customers whose prediction errors in CATE estimation have the greatest impact on profitability, as illustrated in Section 4.1.1.

5.1. Simulation Setup

We consider a firm aiming to develop a targeting policy that maximizes the profitability of a marketing intervention. The firm learns the targeting policy through experimentation on a subset of customers, and subsequently implements the learned policy on future customers (not included in the experiment). We assume that the impact of the (binary) intervention varies across customers.

Building on our problem formulation, we consider two scenarios, illustrated in Figure 1, that reflect distinct ways in which customers can become consequential. In the first scenario, prediction errors are similar across customers. In this case, customers whose treatment effects lie close to the decision threshold — yet slightly deviate from it — are most consequential, as small estimation errors can reverse targeting decisions. In the second scenario, prediction errors vary across customers, such that those with stronger treatment effects are also harder to predict. In this setting, customers far from the threshold can become most consequential because their higher uncertainty increases the likelihood of decision distortions.

We expect our approach to outperform alternative sampling strategies when they fail to prioritize these consequential customers, either when the threshold deviates from where most customers are concentrated (first scenario) or when the most responsive customers are also the hardest to predict (second scenario). To test these predictions, we examine two distinct data-generating processes that reflect these mechanisms. 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 D.6.

Scenario 1: Homoskedastic CATE uncertainty. We first consider a setting where customers have similar prediction errors. We generate a customer base $\mathcal{I}$ and an evaluation set $\mathcal{D}_{\text{eval}}$ 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 SUTVA and Stability assumptions (Assumptions 1 and 2). The resulting CATE distribution is approximately normal with mode and median equal to 1. We vary the intervention cost with $c \in \{0, 1, 2, 3\}$ to examine our approach's performance across different decision environments (see Figure 3 for an illustration). Specifically, $c = 0$ represents scenarios with minimal intervention costs, while $c = 3$ describes situations where only a few customers lie near the decision threshold — common when treatment effects are generally low or intervention costs are prohibitively high (e.g., Ascarza et al. (2016); Lemmens and Gupta (2020)). The case $c = 1$ represents scenarios where customers are predominantly clustered around the decision threshold. Since $c = 0$ and $c = 2$ are positioned symmetrically around the CATE mode (located at 1), this setup allows us to investigate whether performance is influenced solely by the distance from the decision threshold to the mode, or also by the direction relative to the mode.

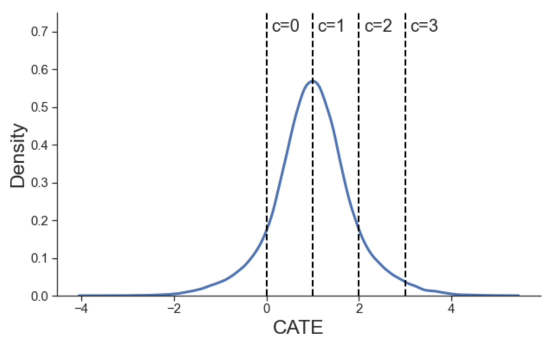


Figure 3: CATE Distribution and Intervention Costs

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.

Scenario 2: Heteroskedastic CATE Uncertainty Positively Correlated with Treatment Effects. We next examine a setting where customers with larger treatment effects also exhibit greater outcome variance that leads to higher CATE uncertainty — a pattern commonly observed when the most responsive customers are also the hardest to predict due to uncertain baseline behavior (e.g., new or infrequent buyers). The data generating process is:

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

where

$$
\begin{aligned}
\tau(X_i) &= X_{i1} \cdot X_{i2} + 0.5 \cdot X_{i3} + 0.6, \\
e_i &\sim \mathcal{N}(0, \sigma_i^2), \quad \sigma_i = 2|\tau(X_i) - 1|, \\
X_{i1} &\sim \text{LogNormal}(-2, 1.5), \\
X_{i2} &\sim \text{Bernoulli}(0.5), \\
X_{i3} &\sim \text{LogNormal}(0, 1), \\
X_{i4} &\sim \text{Bernoulli}(0.5), \\
X_{i5} &\sim \mathcal{N}(0, 1).
\end{aligned}
$$

In this scenario, we fix the intervention cost at $c = 1$ to align the decision threshold with the center of the CATE distribution, allowing us to isolate the role of heteroskedastic uncertainty in determining which customers are most consequential (see Figure 4 for an illustration). Varying the intervention cost in this setting would merely introduce an additional decision-threshold shift, which is already examined in the homoskedastic case.

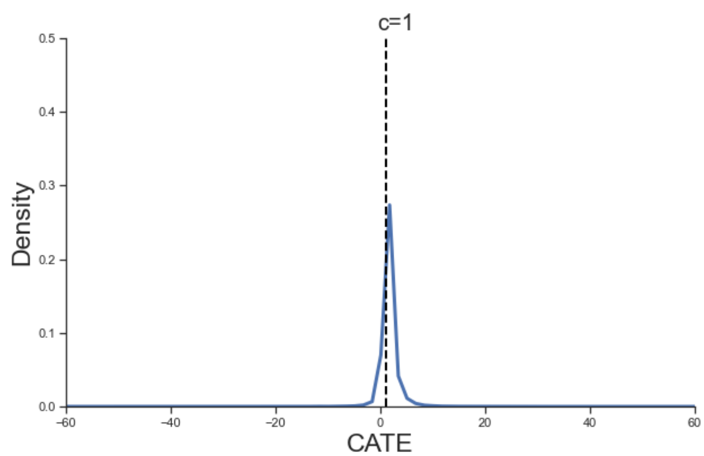


Figure 4: CATE Distribution and Intervention Costs
The dashed line corresponds to a the intervention cost c.

This heteroskedastic error structure $\sigma_i = 2|\tau(X_i) - 1|$ creates a realistic challenge: customers with stronger treatment effects are harder to predict, making their CATE estimates more uncertain. Consequently, high-impact customers are simultaneously the most consequential ones in this scenario. This allows us to assess the performance of different sampling strategies when prediction difficulty is positively correlated with economic impact.

Experimental Size and Design Variants. For both scenarios, we assume a customer base size of $N_{\mathcal{I}} = 100{,}000$. From this population, we draw experimental samples of varying sizes $N_S \in$

$\{3k, 6k, 9k, 12k, 15k\}$ to assess sample efficiency. We evaluate both the multi-stage design (with equal-sized batches of 300 customers) and the simplified two-stage design. In addition, we construct a separate evaluation set of size $N_{\text{eval}} = 10{,}000$ that was not used for the experimentation phase, to assess out-of-sample targeting performance.

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, 300 customers), and another utilizing only two batches of varying size. 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. These settings highlight the potential trade-offs between implementation complexity and targeting efficacy.

5.2.2. Benchmarks

Default. We compare our approach with four 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.

Full. The second benchmark (`Full`) represents an upper bound on performance: conducting the experiment on the entire customer base $\mathcal{I}$ ($N_{\mathcal{S}} = N_{\mathcal{I}} = 100k$) and using all collected data to learn the targeting policy. This allows us to better assess the sample efficiency of our approach.

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

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

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 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.

Mistargeting Prob. The fourth benchmark is the mistargeting-based approach (`Mistargeting Prob`) used in decision-oriented active learning (e.g., Sundin et al. 2019; Filstroff et al. 2021). This approach focuses on sampling customers with the highest probability of being mistargeted under the current model:

$$S_b = \arg\max P\left(\hat{\tau}_{S^{b-1}}(x) \neq \pi^*(x)\right).$$

While both our EPL approach and the mistargeting benchmark account for decision uncertainty, the mistargeting approach ignores the profitability impact of mistargeting, prioritizing only the probability of making an incorrect decision. In contrast, EPL recognizes that firms care about total profitability and explicitly takes that into account. This comparison allows us to empirically assess whether moving beyond decision correctness to a financial objective function can provide additional value in certain contexts.

Adaptive. The fifth benchmark is the state-of-the-art adaptive experimental design in the BAI literature (`Adaptive`) proposed by Kato et al. (2024). We select this as our primary benchmark because it most closely aligns with our research objective: optimizing the exploration phase to learn the most accurate targeting policy for post-experiment deployment. Unlike traditional bandit algorithms that maximize cumulative reward during the experiment by balancing exploration and exploitation, both Kato et al. (2024) and our approach focus purely on improving the quality of the final targeting policy. This shared objective, combined with minimal parametric assumptions, makes it the most appropriate method for comparison.

The key distinction between the two approaches lies in what is being optimized. The Kato et al. (2024) method optimizes *treatment assignment* for sampled customers, assigning customers in batch to different 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. 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. In contrast, our approach optimizes *customer selection* by selectively sampling 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. Comparing these methods allows us to evaluate whether customer selection provides additional value beyond optimal treatment assignment.

5.3. Evaluation Procedure

We evaluate each sampling strategy through a three-phase framework. In the *Experimental phase*, we sample customers from the customer base $\mathcal{I}$ according to the sampling strategy being evaluated (as described in Sections 5.2.1 and 5.2.2). Next, in the *Estimation phase*, we use the collected experimental data to estimate a CATE model. Finally, in the *Rollout phase*, we apply this CATE model to the evaluation set and target the customers, not included in the experimental phase, whose predicted CATE is greater than the intervention cost, which simulates future deployment of policy after the experiment concludes.

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

$$\text{PPG} = \frac{\sum_{i \in D_{eval}} (\tau(\mathbf{X}_i) - c) \cdot \pi^*(\mathbf{X}_i) - \sum_{i \in D_{eval}} (\tau(\mathbf{X}_i) - c) \cdot \hat{\pi}_S(\mathbf{X}_i)}{\sum_{i \in D_{eval}} (\tau(\mathbf{X}_i) - c) \cdot \pi^*(\mathbf{X}_i)} \tag{8}$$

where $\pi^*(\cdot)$ denotes the true optimal policy based on the true CATE $\tau(\cdot)$. The PPG captures the relative profit loss from using the estimated policy instead of the optimal one in the rollout phase. Lower values indicate better performance. (We also investigate the performance of our approach when the final targeting policy is estimated by a direct policy learning method such as Policy Forest (Athey and Wager 2021) in Web Appendix D.2 and Web Appendix E.4 , and find qualitatively similar results, highlighting the generalizability and robustness of our approach across different policy learning methods.)

For each method, we perform 100 replications and present the average across replications.

5.4. Results

We present results separately for homoskedastic and heteroskedastic environments, which correspond to the two theoretical cases discussed in Section 2.

5.4.1. Scenario 1: Homoskedastic CATE Uncertainty

Figure 5 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 ($N_S \in \{3k, 6k, 9k, 12k, 15k\}$). 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 `Default`, `Adaptive`, and `Uncertainty`, 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., $N_S = 3k$ or $6k$), where EPL more effectively concentrates learning on the most consequential customers.

Notably, our approach performs similarly to `Mistargeting Prob` under homoskedastic errors. This similarity arises because both methods prioritize customers near the decision threshold when prediction uncertainty is uniform. Under identical prediction errors, customers closer to the threshold exhibit higher mistargeting probabilities, while high-impact customers far from the threshold are unlikely to be mistargeted regardless of estimation error. We also find that the performance pattern of the multi-stage and the two-stage versions of our approach are similar, with the latter offering similar gains despite its simpler structure. As we note later, in scenario 2, where customer with high responsiveness also have high uncertainty, mistargeting as well as the two stage approach perform worse than the proposed appoach.[10]

The difference in performance between $c = 0$ and $c = 2$ is most attributed to the denominator in Equation (8). To investigate this issue, we also report absolute profit gaps in Figure 6. This

[10]The performance of our approach slightly decreases as sample size grows when $c = 3$. This is likely due to the ability of our approach to quickly identify and sample the limited pool of highly consequential customers in early batches. Once these customers are exhausted, later batches must sample from less consequential customers. This decreases the overall concentration of highly consequential customers in the final experimental dataset, causing the CATE model to distribute attention more broadly rather than focusing intensively on the boundary region. In such cases, firms might benefit from early stopping rules, as discussed in Section 4.3.

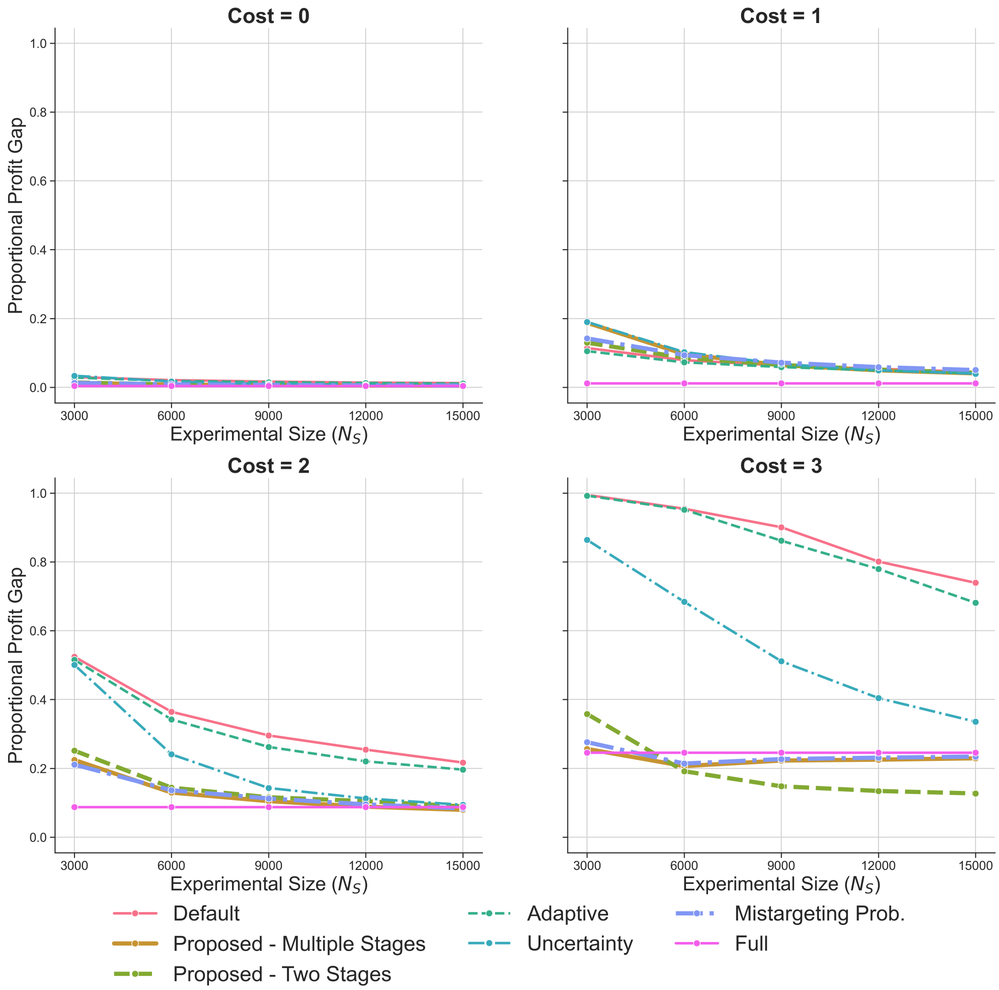


Figure 5: Proportional Profit Gaps under Homoskedastic Errors
We report the average value of the proportional profit gap across 100 replications. Each line corresponds to an experimental approach. Lower values indicate better performance.

comparison confirms that performance is primarily driven by the distance between the decision threshold and the mode of the CATE distribution, as both $c = 0$ and $c = 2$ exhibits similar profit gap in our results.

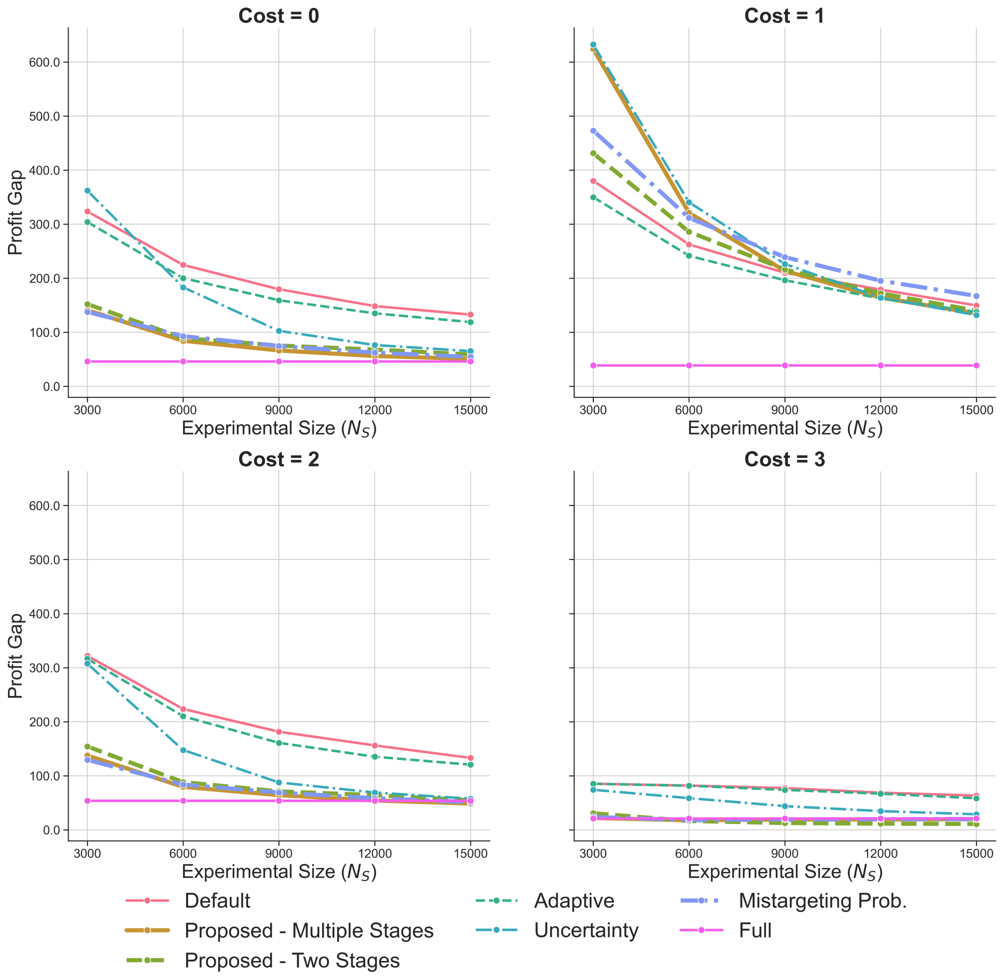


Figure 6: Profit Gaps under Homoskedastic Errors

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.

5.4.2. Scenario 2: Heteroskedastic CATE Uncertainty Positively Correlated with Treatment Effects

We now examine performance of different sampling methods when high-impact customers are also the hardest to predict. Figure 7 shows the proportional profit gaps for intervention cost $c = 1$, positioned at the center of the CATE distribution. (Because the intervention cost is fixed in this scenario, proportional and absolute profit gaps exhibit identical qualitative patterns.)

In contrast to Scenario 1, EPL-based designs now substantially outperform *all* benchmarks across all experimental sizes. Moreover, the multi-stage design outperforms the two-stage design when sample size is small, highlighting the value of iterative refinement of EPL estimates in identifying high-impact customers.

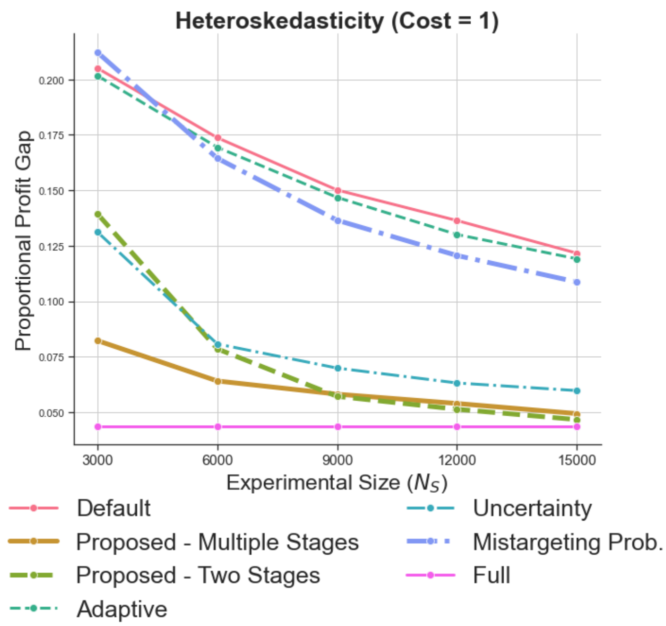


Figure 7: Proportional Profit Gaps Under Heteroskedastic Errors
We report the average value of the proportional profit gap across 100 replications. Each line corresponds to an experimental approach. Lower values indicate better performance.

The key distinction from Scenario 1 emerges in the comparison with `Mistargeting Prob`. In this scenario, customers with the largest treatment effects also exhibit the highest outcome variance, making them not only economically consequential but also likely to be mistargeted when errors occur. EPL explicitly accounts for both factors by weighting mistargeting probability with profit loss magnitude, naturally prioritizing these high-impact, high-uncertainty customers. In contrast, `Mistargeting Prob.` focuses solely on the likelihood of incorrect targeting decisions, allocating substantial sampling effort to customers near the decision threshold who may have modest treatment effects and lower economic stakes. This distinction suggests that when customers with greater economic impact are also harder to predict, methods that ignore profit magnitude can systematically undersample the customers who matter most for targeting profitability.

5.4.3. Discussion

Beyond empirical validation, these results clarify when policy-aware experimentation offers the most value. Under homoskedastic errors where CATE uncertainty is common across customers (Scenario 1), EPL excels when fewer customers lie near the targeting threshold — specifically when intervention costs are misaligned with the center of the CATE distribution. Such scenarios commonly arise when: (1) interventions harm many customers (e.g., Ascarza et al. 2016); (2) interventions are expensive with low effectiveness (e.g., Lemmens and Gupta 2020; Simester et al. 2022); or (3) interventions risk cannibalizing natural revenue (e.g., Anderson and Simester 2004; Ascarza 2018; Yang et al. 2023), which we examine empirically in Section 6.3 (Starbucks).

Under heteroskedastic errors where customers who are most affected by the intervention are also the hardest to predict (Scenario 2), EPL demonstrates broader advantages by outperforming all benchmarks. This superior performance emerges when prediction difficulty correlates with economic impact, which is a realistic pattern where high-value customers (e.g., new or infrequent buyers) are both more responsive and harder to predict, as demonstrated in one of our empirical application in Section 6.2 (Telecom). In such settings, EPL's explicit weighting of both mistargeting probability and profit magnitude proves essential, while methods focusing solely on prediction uncertainty or decision correctness systematically undersample the most consequential customers.

Our approach also exhibits strong sample efficiency. In scenarios where the cost deviates from the mode of the CATE distribution (i.e., $c \in \{0, 2, 3\}$) or harder-to-predict customers are more responsive, our approach matches or exceeds the performance of benchmarks using far fewer observations. Remarkably, we achieve comparable performance to the `Full` benchmark using only 15k samples (15% of the total customer base). In homoskedastic case, when $c = 3$, our approach even slightly outperforms the Full benchmark, demonstrating that strategically focusing on consequential customers can be more effective than broad data collection. Finally, the simplified two-stage design delivers meaningful profit improvement while lagging behind the multi-stage design in several cases. This suggests that two-stage design can be an attractive option to improve targeting performance when the delayed-feedback issue is significant.

5.5. 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.

Scenario 1: Homoskedastic CATE Uncertainty Figure 8 illustrates the CATE distributions of customers sampled by different approaches together with the population distribution. Each subfigure presents one of the scenarios, depending on the intervention cost.[11]

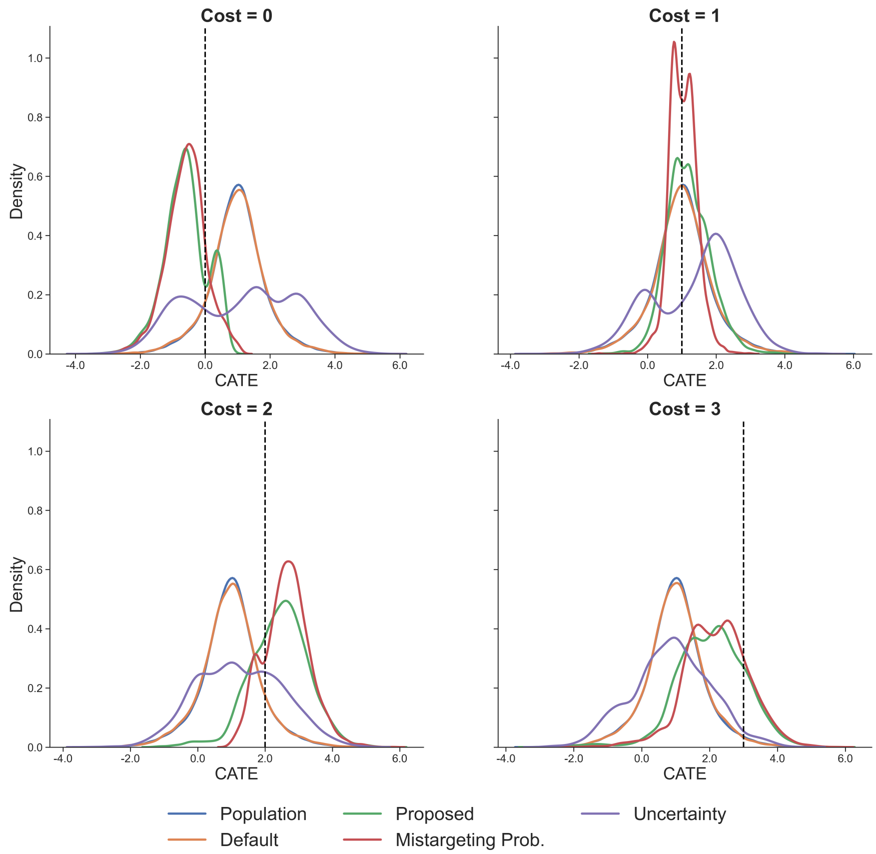


Figure 8: CATE Distributions of Customers Sampled by Different Approaches: Scenario 1
Each line corresponds to the CATE distribution of customers sampled by different approaches. The dashed line represents the intervention cost, which is also the decision threshold.

When CATE uncertainty is homogeneous across customers, both `Proposed` and `Mistargeting Prob` select customers whose CATEs cluster near the decision threshold, explaining their similar performance. When the cost aligns with the mode ($c = 1$), both methods concentrate sampling even more tightly around the threshold than random sampling. When the cost diverges from the mode ($c \in \{0, 2, 3\}$), both methods sample customers close to, though not directly at, the threshold.[12] In contrast, `Uncertainty` produces a more dispersed distribution, oversampling extreme

[11] Here we set the sample size to 10k and implement the EPL sampling strategy using a two-stage design with $r = 0.5$.
[12] When $c = 3$, almost all methods focus on customers that fall below the threshold because limited customers exist above the threshold. For $c \in \{0, 2\}$, methods intensify sampling in regions underrepresented by initial random sampling.

values at the expense of economically relevant boundary regions, while `Default` mirrors the population distribution.

Scenario 2: Heteroskedastic CATE Uncertainty Positively Correlated with Treatment Effects Figure 9 reveals a difference in sampling patterns when prediction difficulty correlates with economic impact. `Mistargeting Prob` concentrates heavily around the threshold ($c = 1$), similar to Scenario 1, as it prioritizes customers with high mistargeting likelihood regardless of outcome variance. In contrast, `Proposed` and `Uncertainty` both spread sampling more broadly toward high-CATE regions where outcome variance is elevated. This distinction explains why EPL substantially outperforms `Mistargeting Prob`, and highlights the importance of accounting for consumer's economic impact in addition to decision correctness when consumers more responsive to the intervention are also less predictable.

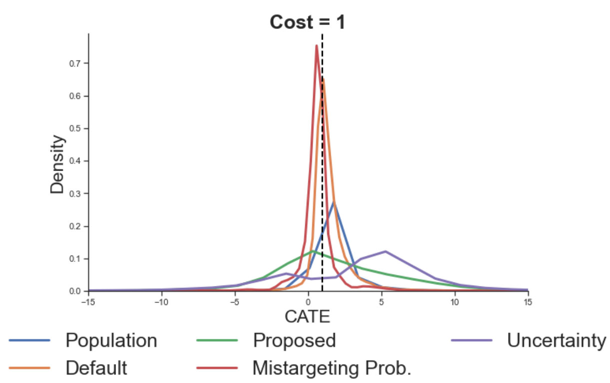


Figure 9: CATE Distributions of Customers Sampled by Different Approaches: Scenario 2
Each line corresponds to the CATE distribution of customers sampled by different approaches. The dashed line represents the intervention cost $c = 1$. The distribution is truncated at $-15, 15$.

These results demonstrate that our approach effectively identifies and prioritizes consequential customers across both scenarios, with the distinction from alternative methods most pronounced when either (1) The mass of customers lies away from the decision threshold (Scenario 1), or (2) prediction difficulty positively correlates with economic impact (Scenario 2). To further validate that our EPL estimates correctly identify these consequential customers, we conduct Spearman rank correlation tests between estimated and true EPLs. Across all scenarios and sample sizes, we find statistically significant positive correlations ($\rho > 0.6$, $p < 0.001$). Additionally, customers sampled using our EPL estimates overlap with approximately 60% of customers with truly highest EPLs. Together, these findings confirm that our method successfully identifies economically consequential customers despite not observing true CATEs.

In conclusion, our simulation results affirm the superiority of our approach in enhancing targeting performance and sample efficiency across different settings. Under homoskedastic CATE uncertainty, EPL excels when few customers lie near the decision threshold. Under heteroskedastic errors, EPL provides broader advantages by correctly weighting both prediction uncertainty and economic impact, outperforming methods that focus solely on mistargeting probability. Taken together, these findings validate the theoretical foundations of our method and demonstrate its practical value: firms can achieve substantial gains in targeting precision and experimental efficiency by aligning sampling strategies with profit-maximizing objectives, particularly when consequential customers are either scarce or difficult to predict. 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. These empirical applications mirror the two simulation scenarios studied in Section 5, allowing us to assess whether the mechanisms identified in controlled settings extend to real-world data.[13] 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 `Full` targeting policy learned from the entire customer base, the `Uncertainty` targeting policy learned from the observed outcomes of the customers sampled by the uncertainty approach, the `Mistargeting Prob` targeting policy learned from the observed outcomes of the customers sampled by the mistargeting approach, and the `Proposed` policy, which is learned by observing the outcomes of the customers selected via our proposed sampling strategy.[14] Consistent with the analyses in previous sections, we evaluate different variants of our

[13] We were not involved in designing or running these original experiments; rather, we use the existing experimental data to evaluate different sampling strategies retrospectively

[14] 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

approach by modifying the number of stages: from multiple batches (each with a size of 500) to a simplified two-stage design.

To assess the performance of each targeting approach, we use a 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%. For each split, we follow a three-phase procedure analogous to Section 5. In the Experimental Phase, we simulate different sampling strategies by selecting subsets of customers from $\mathcal{I}$ according to each sampling criterion (`Default`, `Uncertainty`, `Mistargeting Prob`, and `Proposed`), using the actual treatment assignments and observed outcomes from the original experiment. These sampled customers constitute the experimental data S and are used to estimate a CATE model for each approach in the Estimation Phase. Finally, in the Rollout Phase, we leverage the constructed CATE model to generate targeting decisions $\hat{\pi}(\cdot)$ on the evaluation set, D_{eval}, and evaluate their performance.

We evaluate the targeting performance by first computing the expected profit generated by each estimated targeting policy, $\hat{\pi}_S(\cdot)$, as well as the optimal uniform policy indicated by the average treatment effect, π_u^*. In particular, the optimal uniform policy treats everyone if the ATE is positive and treats no one if the ATE is negative. This represents the best policy that does not utilize customer characteristics for targeting.

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, $\hat{\pi}(\cdot)$:

$$\text{Profit}(\hat{\pi}) = \sum_i \left(\frac{1 - W_i}{1 - e(\mathbf{X}_i)} (1 - \hat{\pi}_S(\mathbf{X}_i)) Y_i(0) + \frac{W_i}{e(\mathbf{X}_i)} \hat{\pi}_S(\mathbf{X}_i)(Y_i(1) - c_i) \right) \tag{9}$$

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 optimal uniform policy π_u^* is then computed as the evaluation metric using the following formula:

$$\text{PPI} = \frac{\text{Profit}(\hat{\pi}_S) - \text{Profit}(\pi_{u^*})}{\text{Profit}(\pi_{u^*})} \tag{10}$$

where $\text{Profit}(\pi_u^*)$ denotes the expected profit of the optimal uniform policy. Since this metric quantifies the profit improvement of the estimated targeting policies relative to the optimal uni-

low response rates in both datasets, accurately predicting customers' counterfactual behavior in these contexts is very challenging.

form policy, a larger PPI indicates that the targeting decision estimated from the experimental data D_e is more profitable.

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 (Ebbes et al. 2026). 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, with an average treatment effect of -0.65, corresponding to a 5.8% decrease in average recharge.

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 E.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 from our simulation study (Section 5) that the proposed approach should be particularly useful when more unpredictable customers are also more responsive to treatment. Our data exhibits this scenario: we observe a correlation of 0.74 between $|\hat{\tau}(x) - c|$ and the standard error of treatment effect estimates, indicating that customers whose treatment effects are more uncertain are indeed more responsive to the intervention.

6.2.2. Profitability of Targeting Policies

We now turn to evaluate the performance of different policies. To ensure clean out-of-sample validation, we split the data into customer base, $\mathcal{I}$, used to run each experimental approach, and evaluation set, D_{eval}, used to evaluate the profitability of each policy, as defined in Equation (10). Figure 10 shows the proportional profit improvement of the targeting policies relative to the uniform policy on the evaluation set. 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 different approaches.[15]

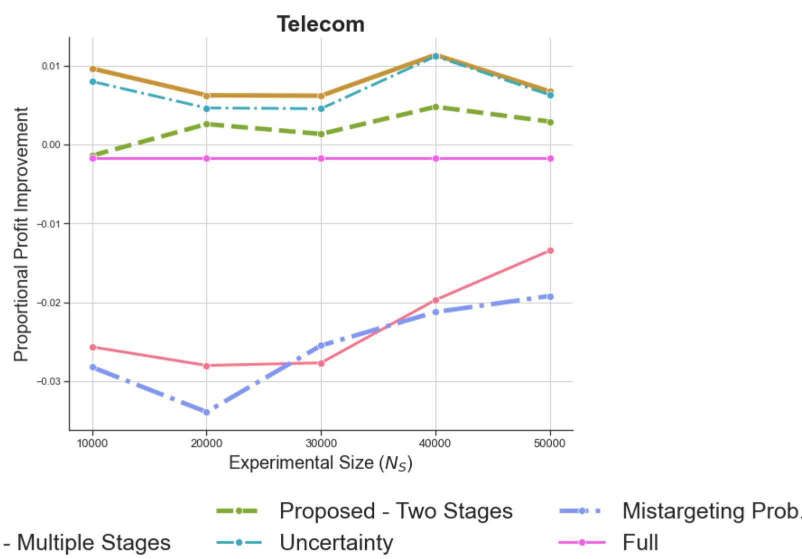


Figure 10: Performance of Targeting Policies Learned from Different Experimental Designs (Telecommunication)

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 yields higher profit improvement than the `Default` and `Mistargeting Prob` approach and performs comparably to the `Uncertainty` approach.[16] This performance pattern aligns with our simulation findings and the data characteristics: when customer with high responsiveness to the treatment are more difficult to predict, it

[15] We do not report standard errors for the IPW estimator due to its large variance, particularly when the evaluation policy deviates from the behavior policy (Saito et al. 2021; Bhargava et al. 2024).

[16] 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)

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.

Experimental Design	Experimental Size				
	10k	20k	30k	40k	50k
Proposed Multiple Stages	70%	69%	67%	73%	68%
Proposed Two-Stage	62%	68%	68%	69%	59%
Uncertainty Sampling	66%	69%	66%	69%	70%
Mistargeting Prob	49%	51%	46%	52%	48%

is critical to consider both the economic impact and decision correctness, rather than focusing on decision correctness alone.

Remarkably, our approach achieves higher profit improvement than the Full benchmark, which experiments on the entire customer base of approximately 300k customers, across all sample sizes. This counterintuitive result reveals an important insight: in noisy data environments, more data is not always better. When noise is high, additional observations carry limited signal and are unlikely to meaningfully improve estimation accuracy, while diluting the concentration of consequential customers in the training data. This dilution causes the model to distribute attention across all customers rather than focusing intensively on those where targeting decisions are most uncertain and economically impactful, and can be particularly pronounced when using local estimators that rely on the similarity between training and target observations (e.g. Wager and Athey 2018; Athey et al. 2019b).[17]

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. 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

[17] This pattern mirrors our simulation findings with $c = 3$ (Section 5.4.1), where performance can degrade with additional broad sampling once consequential customers are exhausted.

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.[18] See Web Appendix E.2 for further details about the data.

Unlike the telecommunication dataset, this data exhibits a low correlation of only 0.13 between the profitability of treatment, $|\hat{\tau}(x) - c|$, and the standard error of treatment effect estimates. This near-homoskedastic pattern suggests that prediction uncertainty is relatively uniform across customers, reducing the importance of balancing economic impact and decision correctness in the sampling strategy.

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. 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.[19]. 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.

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 (10). Figure 11 shows the proportional profit improvement of the targeting policies relative to the uniform policy for each scenario (no discount and 50%-off discount).

[18]The data were provided by Starbucks through the Udacity Data Science Program. The dataset is anonymized and does not include descriptions of the pre-treatment variables. Although this limits interpretability, the variables contain sufficient information to construct targeting policies, which is the primary objective of our analysis.

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

Each line corresponds to the profitability of the policy derived by each experimental design (or sampling approach), across experimental sizes ($N_S \in \{30k, 35k, 40k, 45k, 50k\}$).

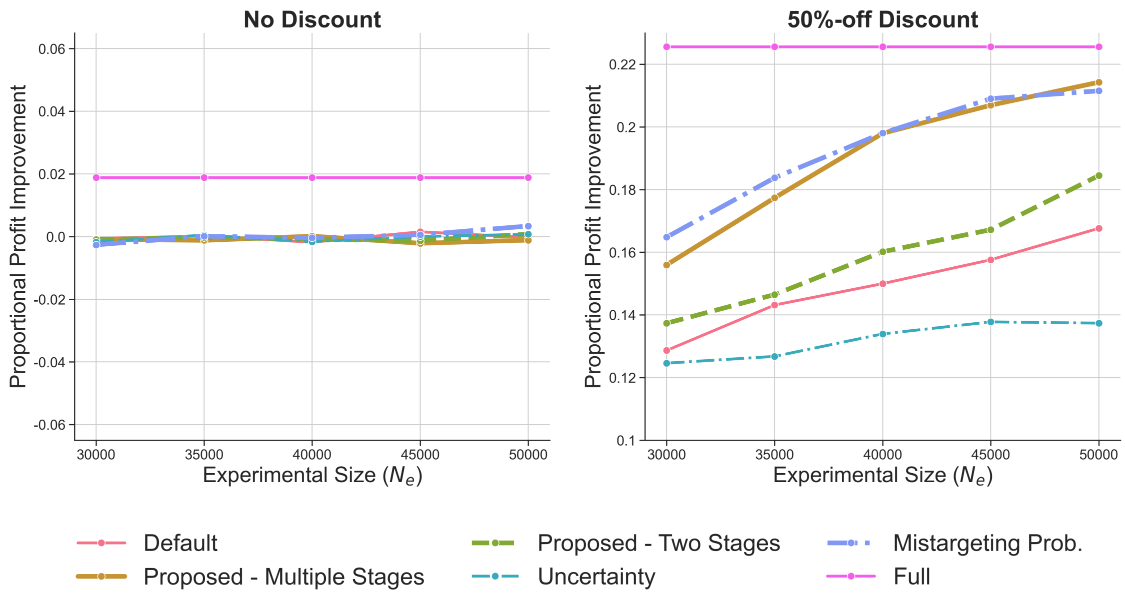


Figure 11: Performance of Targeting Policies Learned from Different Experimental Designs (Starbucks)
We report the average value of the proportional profit improvement relative to the uniform policy across 100 replications. Each line corresponds to an experimental approach. Higher values indicate better performance.

First, our approach significantly outperforms both `Default` and `Uncertainty` 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 all benchmark approaches 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.

Notably, the mistargeting probability approach performs comparably to our proposed method in this application, contrasting with the telecommunication case where our approach substantially outperformed it. This difference can be attributed to the minimal heteroskedasticity in the Starbucks data, which reduces the advantage of jointly considering economic impact and decision correctness. When heteroskedasticity is low, sampling based primarily on decision correctness (`Mistargeting Prob`) or jointly considering both factors (our approach) yields similar results, as high-impact consumers are unlikely to be mistargeted. Second, on the flip side, unlike telecommunication application, the `Uncertainty` benchmark performs poorly in this application, reflecting the homoskedastic nature of this setting where prediction difficulty does not correlate with treatment effect magnitude, and where the high intervention cost makes decision correctness the primary driver of targeting performance.[20]

Table 2: Percentage of Replications in which the Proposed Method Outperforms the Default Approach (Starbucks)

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.

Experimental Design	Experimental Size				
	30k	35k	40k	45k	50k
Proposed Multiple Stages	65%	75%	84%	84%	88%
Proposed Two-Stage	57%	49%	57%	53%	74%
Uncertainty Sampling	50%	38%	34%	32%	25%
Mistargeting Prob	71%	75%	79%	85%	86%

Third, even when the proposed approach does not provide additional benefits in learning targeting policies, as in no discount case, 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.

Fourth, 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 three benchmarks. The magnitude of performance gains, however, depends on two key data characteristics identified in our simulation study: (1) the degree of misalignment between the CATE distribution and the decision threshold, and (2) the relationship between the CATE uncertainty and CATE estimates.

First, our method excels when the intervention creates a risk of cannibalizing profits that would have been earned in the absence of intervention, as evidenced in the 50% discount scenario

[20] Unlike the telecommunication application, our approach underperforms the Full benchmark here. In less noisy settings, additional observations carry more signal and improve estimation accuracy, making broad data collection more competitive with strategic sampling.

for Starbucks (offering discounts). In such cases, the interventions reduce incremental 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). Thus, we recommend particularly adopting the proposed approach, particularly when there is a likely discrepancy between intervention cost and customer responsiveness.[21]

The telecommunication campaign highlights another scenario in which the proposed approach shines: when customers with more unpredictable behavior are also more responsive to the intervention. Under these conditions, our method substantially outperformed the mistargeting probability approach. In contrast, the Starbucks application exhibited minimal heteroskedasticity (correlation of 0.13), resulting in comparable performance between both methods. This demonstrates that when heteroskedasticity is present, jointly considering economic impact and decision correctness provides meaningful benefits. Such heteroskedastic patterns may arise in several business scenarios where customers with less stable behavior, such as infrequent buyers or new customers, exhibit higher responsiveness to interventions. Taken together, since firms rarely know which of these conditions characterizes their setting ex ante, EPL offers a robust and attractive default strategy delivering strong performance when either condition holds, while alternative methods tend to perform well only in certain scenarios.

Finally, across both applications, the simplified two-stage design delivers profit improvement relative to the `Default` approach while lagging behind the multi-stage approach. This suggests

[21] 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.

that firms may incur a trade-off between profit gains and simplicity, especially when delayed-feedback is pronounced.

7. Conclusion

This paper develops a sampling strategy tailored to profit-maximizing targeting decisions, *policy-aware sampling*, 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 an *expected profit loss (EPL)* criterion that quantifies customers' learning value based on the impact of estimation error on targeting profitability, and show that sampling based on this criterion yields a near-optimal approximation to the sampling strategy that maximizes the profitability of the resulting 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 theoretical properties of Causal Forest estimators. This enables firms to identify and oversample "consequential" customers without requiring knowledge of true treatment effects. The method can be implemented in either a multi-stage 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 sampling offers substantial improvements over conventional sampling strategies — including standard test-then-learn, uncertainty sampling, and a leading adaptive strategy — especially in contexts where customers with unpredictable behavior (infrequent buyers, new buyers, etc.) are likely to be more responsive to the intervention, or 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. In contrast, when the intervention is costless and broadly beneficial, a simple strategy that targets all customers is already close to optimal, leaving limited room for further improvement in targeting performance, as shown in Section 6.3.2.

Managerial Implications. For marketing leaders aiming to improve the profitability of experimental targeting, our findings offer actionable guidance. The expected profit loss (EPL) sampling strategy is particularly valuable when customers with unpredictable behavior (infrequent buyers, new buyers, etc.) are likely to be more responsive to the intervention, or 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 97% fewer experimental observations. These results underscore our approach efficiency in data usage and impact on decision quality, making it 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. When delayed feedback makes fully adaptive designs impractical, a simplified two-stage implementation can also deliver meaningful benefits while minimizing operational burden, though firms should be aware that some profit gains may diminish as a trade-off.

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' targeting objectives are not subject to any constraints. However, in practice, firms may face managerial constraints on their targeted policies, such as budget constraints or fairness constraints (Lu et al. 2023). Our framework can partially accommodate budget constraints by introducing an elevated decision threshold that limits the number of customers treated, which we discuss in Web Appendix F.2. However, a fully optimal treatment of budget constraints would require the endogenous determination of this threshold under uncertainty, accounting for the fact that the optimal threshold itself depends on the distribution of treatment effects that has yet to be fully learned. 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 accounts for these constraints.

Second, 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 relationship-based marketing applications, 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. We discuss a preliminary extension of our framework to this setting in Web Appendix F.4, though a fully optimal solution remains an open methodological challenge. 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, while our framework focuses on binary treatment assignments consistent with our empirical applications, the EPL criterion is a general policy-aware value-of-information principle that extends naturally to multiple treatment arms. We formalize this extension in Web Appendix F.3, though the complete modeling approach and empirical evaluation across multi-arm settings remains a promising avenue for future research.

Overall, our research demonstrates the value of incorporating firms' business objectives into the design of experimental sampling strategy. 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

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.

Bhargava, A., Jain, L., Kveton, B., Liu, G., and Mukherjee, S. (2024). Off-Policy Evaluation from Logged Human Feedback. Version Number: 1.

Blattberg, R. C. (1979). The Design of Advertising Experiments Using Statistical Decision Theory. *Journal of Marketing Research*, 16(2):191.

Bradlow, E. T. and Wainer, H. (1998). Some statistical and logical considerations when rescoring tests. *Statistica Sinica*, 8(3):713–728.

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.

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.

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].

Ebbes, P., Ascarza, E., and Netzer, O. (2026). Learning from many experiments: A hierarchical bayesian framework for decomposing marketing treatment heterogeneity. Working paper, Working Paper.

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.

Elmachtoub, A. N. and Grigas, P. (2022). Smart "Predict, then Optimize". *Management Science*, 68(1):9–26.

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.

Filstroff, L., Sundin, I., Mikkola, P., Tiulpin, A., Kylmäoja, J., and Kaski, S. (2021). Targeted Active Learning for Bayesian Decision-Making. *arXiv preprint arXiv:2106.04193*. arXiv:2106.04193 [cs, stat].

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.

Ginter, J. L., Cooper, M. C., Obermiller, C., and Page, T. J. (1981). The Design of Advertising Experiments Using Statistical Decision Theory: An Extension. *Journal of Marketing Research*, 18(1):120–123.

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.

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

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*.

Ibragimov, M., Simester, D., and Timoshenko, A. (2025). Improving Targeting Policies Using Transfer Learning.

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.

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].

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.

Ryzhov, I. O., Powell, W. B., and Frazier, P. I. (2012). The Knowledge Gradient Algorithm for a General Class of Online Learning Problems. *Operations Research*, 60(1):180–195.

Saito, Y., Udagawa, T., Kiyohara, H., Mogi, K., Narita, Y., and Tateno, K. (2021). Evaluating the Robustness of Off-Policy Evaluation. Publisher: arXiv Version Number: 1.

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].

Toubia, O. and Hauser, J. R. (2007). **Research Note** —On Managerially Efficient Experimental Designs. *Marketing Science*, 26(6):851–858.

Urban, G. L., Liberali, G. G., MacDonald, E., Bordley, R., and Hauser, J. R. (2014). Morphing Banner Advertising. *Marketing Science*, 33(1):27–46.

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.

Xie, M. and Singh, K. (2013). Confidence Distribution, the Frequentist Distribution Estimator of a Parameter: A Review. *International Statistical Review*, 81(1):3–39.

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 Sampling: Prioritizing Consequential Customers 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

Method	Objective	Sampling Strategy	Sampling Space	Pre-treatment Covariates	Pre-Segmentation	Sequential	References
Managerially Efficient Experimental Design	Improve managerial efficiency	Optimize M-error	Context	N	N	N	Toubia and Hauser (2007)
Test and Roll	Calculate sample size	N/A	N/A	N	N	N	Feit and Berman (2019)
Sample Size Computation	Calculate sample size	N/A	N/A	Y	Y	N	Simester et al. (2022)
Minimax Sample Selection	Learning minimax-regret policy	Stratified Sampling	Context	Y	Y	N	Hu et al. (2024)
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), Jain et al. (2024)
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)
Online Causal Inference	Reducing experimentation Cost for ad effect estimation	TS	Action	Y	N	Y	Waisman et al. (2024)
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)
Decision-aware Learning	Aligning prediction with downstream decision-making	N/A	N/A	Y	N	N	Elmachtoub and Grigas (2022), Chung et al. (2022)
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 Lipschitz continuity assumption is a mild smoothness condition that simply requires customers with similar characteristics to have similar treatment effects. This is a standard assumption in the nonparametric treatment effect literature and explicitly adopted by Wager and Athey (2018) and Athey et al. (2019) in establishing the theoretical properties of causal forests. 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,

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

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

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

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. Derivation of EPL Function under Normal Confidence Distribution

For an asymptotically normal estimator $\hat{\tau}_{\mathcal{D}}(\mathbf{X}_i)$ estimated from a dataset $\mathcal{D}$, the confidence distribution of the parameter $\tau(\mathbf{X}_i)$ can be expressed as:

$$h_{\mathcal{D}}(\tau(\mathbf{X}_i)) = \frac{1}{\hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)}\phi\left(\frac{\tau(\mathbf{X}_i) - \hat{\tau}_{\mathcal{D}}(\mathbf{X}_i)}{\hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)}\right)$$

where ϕ is the density function of the standard normal distribution and $\hat{\sigma}(\mathbf{X}_i)$ is the consistent estimator of the standard error $\sigma(\mathbf{X}_i)$.

Since confidence distribution can be treated as a distribution estimator of $\tau(\mathbf{X}_i)$ in the style of a Bayesian posterior (Xie and Singh 2013), the EPL for customer i can be written as the expected profit loss from potential deviations between the estimated and potential optimal targeting decisions, integrated over the confidence distribution:

$$\begin{aligned}
\hat{\ell}_{\mathcal{D}}(\mathbf{X}_i) &= \int \underbrace{0 \cdot \mathbf{1}\{\hat{\pi}_{\mathcal{D}}(\mathbf{X}_i) = \pi(\mathbf{X}_i)\}}_{\text{profit loss with no deviation}} + \underbrace{|\pi(\mathbf{X}_i) - c(\mathbf{X}_i)|\mathbf{1}\{\hat{\pi}_{\mathcal{D}}(\mathbf{X}_i) \neq \pi(\mathbf{X}_i)\}}_{\text{profit loss with deviation}} h_{\mathcal{D}}(\tau(\mathbf{X}_i))d\tau \\
&= \int \underbrace{|\tau(\mathbf{X}_i) - c(\mathbf{X}_i)|\mathbf{1}\{\hat{\pi}_{\mathcal{D}}(\mathbf{X}_i) \neq \pi(\mathbf{X}_i)\}}_{\text{profit loss}} h_{\mathcal{D}}(\tau(\mathbf{X}_i))d\tau \\
&= \int \underbrace{|\tau(\mathbf{X}_i) - c(\mathbf{X}_i)|\mathbf{1}\{\hat{\pi}_{\mathcal{D}}(\mathbf{X}_i) \neq \pi(\mathbf{X}_i)\}}_{\text{profit loss}} \frac{1}{\hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)}\phi\left(\frac{\tau(\mathbf{X}_i) - \hat{\tau}_{\mathcal{D}}(\mathbf{X}_i)}{\hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)}\right) d\tau
\end{aligned} \tag{W-3}$$

If $\hat{\tau}_{\mathcal{D}}(\mathbf{X}_i) \geq 0$, we have

$$\begin{aligned}
\hat{\ell}_{\mathcal{D}}(\mathbf{X}_i) &= \int |\tau(\mathbf{X}_i) - c(\mathbf{X}_i)|\mathbf{1}\{\hat{\pi}_{\mathcal{D}}(\mathbf{X}_i) \neq \pi(\mathbf{X}_i)\} \frac{1}{\hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)}\phi\left(\frac{\tau(\mathbf{X}_i) - \hat{\tau}_{\mathcal{D}}(\mathbf{X}_i)}{\hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)}\right) d\tau \\
&= \int_{-\infty}^{c(\mathbf{X}_i)} (c(\mathbf{X}_i) - \tau(\mathbf{X}_i)) \cdot \frac{1}{\hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)}\phi\left(\frac{\tau(\mathbf{X}_i) - \hat{\tau}_{\mathcal{D}}(\mathbf{X}_i)}{\hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)}\right) d\tau.
\end{aligned} \tag{W-4}$$

Let

$$z = \frac{\tau(\mathbf{X}_i) - \hat{\tau}_{\mathcal{D}}(\mathbf{X}_i)}{\hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)}.$$

Since $\tau(\mathbf{X}_i) = \hat{\tau}_{\mathcal{D}}(\mathbf{X}_i) + rz\hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)$ and $d\tau = \hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)dz$, we have

$$\begin{aligned}
\hat{\ell}_{\mathcal{D}}(\mathbf{X}_i) &= \int_{-\infty}^{c(\mathbf{X}_i)} (c(\mathbf{X}_i) - \tau(\mathbf{X}_i)) \cdot \frac{1}{\hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)}\phi\left(\frac{\tau(\mathbf{X}_i) - \hat{\tau}_{\mathcal{D}}(\mathbf{X}_i)}{\hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)}\right) d\tau \\
&= \int_{-\infty}^{-r(\mathbf{X}_i)} (c(\mathbf{X}_i) - \hat{\tau}_{\mathcal{D}}(\mathbf{X}_i) - z\hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i))\phi\left(z\right))\, dz \\
&= (c(\mathbf{X}_i) - \hat{\tau}_{\mathcal{D}}(\mathbf{X}_i)) \int_{-\infty}^{-r(\mathbf{X}_i)} \phi(z)dz - \hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i) \int_{-\infty}^{-r(\mathbf{X}_i)} z\phi(z)dz
\end{aligned}$$

where

$$r(\mathbf{X}_i) = \frac{|\hat{\tau}_{\mathcal{D}}(\mathbf{X}_i) - c(\mathbf{X}_i)|}{\hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)}.$$

Since $\int_{-\infty}^{a} \phi(z)dz = \Phi(a)$ and $\int z\phi(z)dz = -\phi(z)$, we have

$$\begin{aligned}
\hat{\ell}_{\mathcal{D}}(\mathbf{X}_i) &= (c(\mathbf{X}_i) - \hat{\tau}_{\mathcal{D}}(\mathbf{X}_i))\Phi(-r(\mathbf{X}_i)) - \hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)(-\phi(-r(\mathbf{X}_i))) \\
&= -\hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)r(\mathbf{X}_i)\Phi(-r(\mathbf{X}_i)) + \hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)\phi(r(\mathbf{X}_i)) \\
&= \hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)(\phi(r(\mathbf{X}_i)) - r(\mathbf{X}_i)\Phi(-r(\mathbf{X}_i))).
\end{aligned}$$

Similarly, if $\hat{\tau}_{\mathcal{D}}(\mathbf{X}_i) < 0$,

$$\hat{\ell}_{\mathcal{D}}(\mathbf{X}_i) = \hat{\sigma}_{\mathcal{D}}(\mathbf{X}_i)(\phi(r(\mathbf{X}_i)) - r(\mathbf{X}_i)\Phi(-r(\mathbf{X}_i))).$$

Therefore, we have

$$\hat{\ell}_{\mathcal{D}}(\mathbf{X}_i) = \hat{\sigma}_{\mathcal{D}}\,(\mathbf{X}_i)(\phi(r(\mathbf{X}_i)) - r(\mathbf{X}_i)\Phi(-r(\mathbf{X}_i)))$$

.

D. Further Details of Simulation Studies

D.1. Implementation Details

D.1.1. Proposed Approach with Multiple Stages

For each experimental size $N_S \in \{3k, 6k, 9k, 12k, 15k\}$, 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 $\frac{N_S}{300}$ batches, each containing 300 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, chosen based on cross-validation.

D.1.2. Proposed Approach with Two Stages

For each experimental size $N_S \in \{3k, 6k, 9k, 12k, 15k\}$, 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, chosen based on cross-validation.

D.1.3. Default (A/B Test with Random Sampling)

For each experimental size $N_S \in \{3k, 6k, 9k, 12k, 15k\}$, 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, chosen based on cross-validation.

D.1.4. Full (A/B Test with Entire Customer Base)

We follow the same procedure as the `Default` approach but sampling the entire customer base $\mathcal{I}$.

D.1.5. Uncertainty Sampling

For each experimental size $N_S \in \{3k, 6k, 9k, 12k, 15k\}$, 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 $\frac{N_S}{300}$ batches, each containing 300 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, chosen based on cross-validation. The uncertainty (standard error) of the CATE estimates $\sigma(\hat{\tau}_{S^{b-1}}(x))$ is estimated using the bootstrap-of-little-bag approach (Athey et al. 2019) implemented in the package.

D.1.6. Mistargeting Probability

We follow the same procedure as for the `Uncertainty` approach, but sample the customers with the largest mistargeting probability. For mistargeting probability 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, chosen based on cross-validation. The mistargeting probability is estimated by the formula $\Pr\left(\hat{\pi}S^{b-1}(x) \neq \pi^*(x)\right) = \Phi\left(\frac{-|\hat{\tau}_{S^{b-1}}(x)-c|}{\hat{\sigma}_{S^{b-1}}(x)}\right)$ where $\hat{\sigma}_{S^{b-1}}(x)$ is estimated using the bootstrap-of-little-bag approach (Athey et al. 2019) implemented in the package.

D.1.7. Adaptive (Kato et al. 2024)

For each experimental size $N_S \in \{3k, 6k, 9k, 12k, 15k\}$, 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 $\frac{N_S}{300}$ batches, and randomly sample 300 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.[1]

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, chosen based on cross-validation. Note that the Causal Forest model in `econML` is designed to solve the local moment equation:

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

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

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):

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

D.2. Additional Results for Direct Policy Learning

In this appendix, we present additional results examining the performance of our approach when the final targeting policy is estimated using a direct policy learning method. Specifically, we implement Policy Forest (Athey and Wager 2021), which directly learns the optimal policy by maximizing expected outcomes without explicitly estimating conditional average treatment effects. This analysis assesses whether our method's advantages persist when paired with policy learning methods that optimize decision rules directly rather than through predicted treatment effects, and thus being generalizable beyond Causal Forest.[2]

Figure W-1 and Figure W-2 show the proportional profit gaps of targeting policies learned by different experimental designs under alternative sample allocation schemes across intervention costs ($c \in \{0, 1, 2, 3\}$ for Scenario 1; $c = 1$ for Scenario 2) and experimental sizes ($N_S \in \{3k, 6k, 9k, 12k, 15k\}$). The results are qualitatively similar to those in Section 5.4. Under homoskedastic errors, our approach performs comparably to `Mistargeting Prob.` and outperforms other benchmarks when few customers lie near the threshold ($c \in \{0, 2, 3\}$), especially at smaller sample sizes. Under heteroskedastic errors, EPL outperforms all benchmarks including `Mistargeting Prob.`, especially when sample size is small. These results demonstrates that our method's advantages persist regardless of whether policies are learned through CATE estimation or direct policy optimization. These findings confirm that EPL-based sampling effectively improves targeting performance across different policy learning approaches by strategically prioritizing consequential customers.

[2] We implement the Policy Forest method using the `econML` package in Python. The model comprises 100 trees with maximum depths not exceeding 10, chosen based on cross-validation.

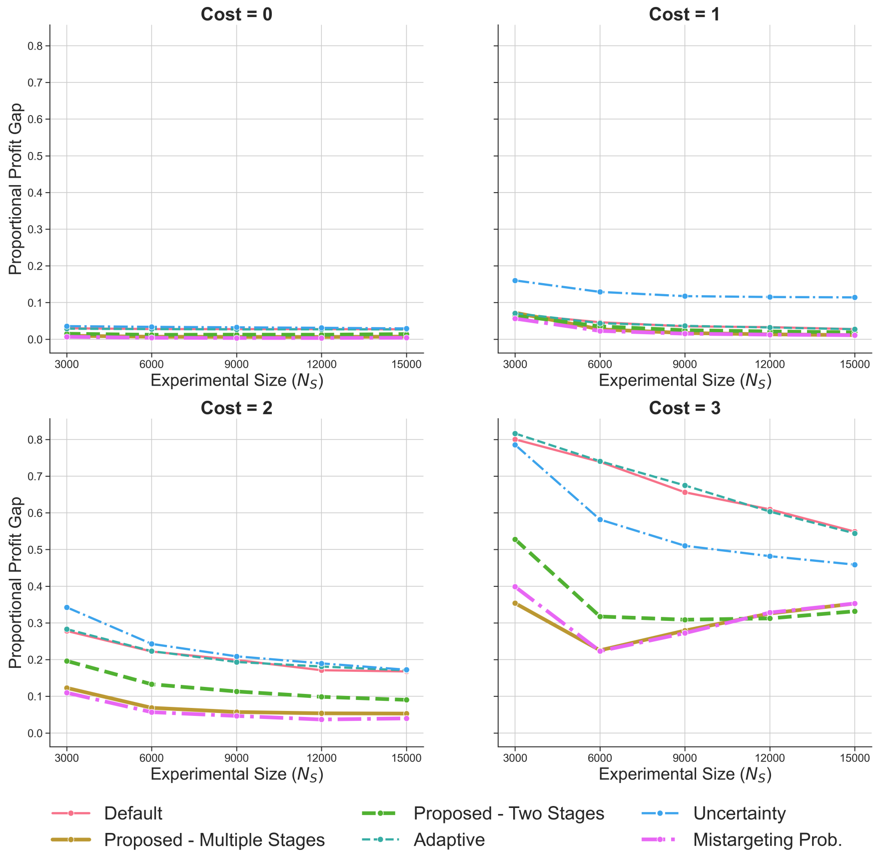


Figure W-1: Proportional Profit Gaps under Homoskedastic Errors with Direct Policy Learning
We report the average value of the proportional profit gap across 100 replications. Each line corresponds to an experimental approach.

D.3. 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.

D.3.1. 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

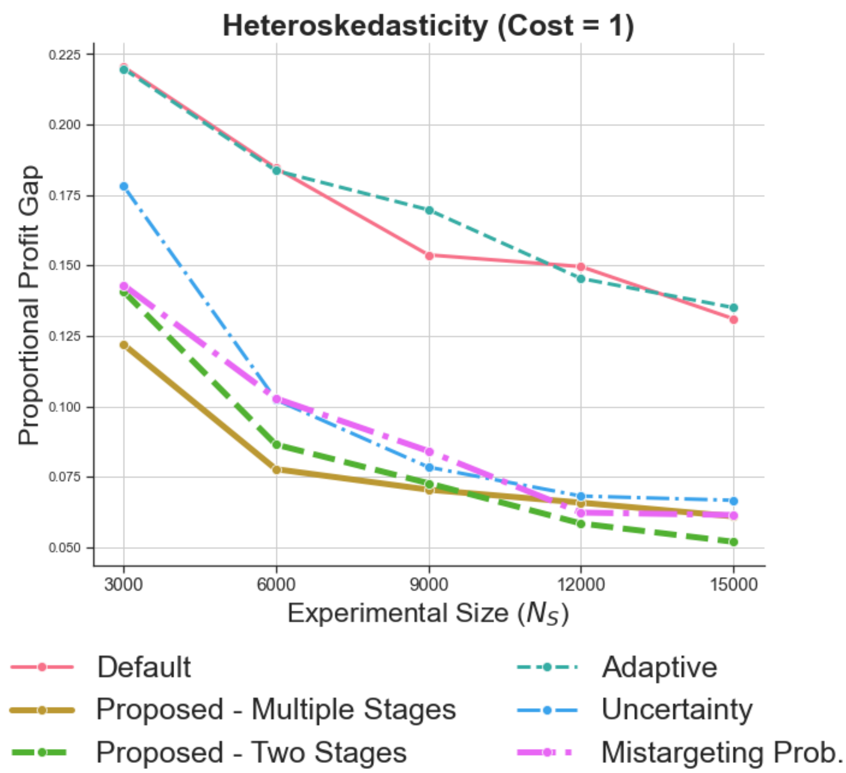


Figure W-2: Proportional Profit Gaps under Heteroskedastic Errors with Direct Policy Learning

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

equally across 10 batches) with a decreasing batch size allocation (where earlier batches receive more samples).

Figures W-3 and W-4 show the proportional profit gaps of targeting policies learned by different experimental designs under alternative sample allocation schemes across intervention costs ($c \in \{0, 1, 2, 3\}$ for Scenario 1; $c = 1$ for Scenario 2) and experimental sizes ($N_S \in \{3k, 6k, 9k, 12k, 15k\}$). The results are qualitatively similar to those with the baseline fixed batch size of 300. Under homoskedastic errors (Figure W-3), our approach performs comparably to `Mistargeting Prob.` and outperforms other benchmarks when few customers lie near the threshold ($c \in \{0, 2, 3\}$), especially at smaller sample sizes. Under heteroskedastic errors (Figure W-4), EPL substantially outperforms all benchmarks including `Mistargeting Prob.` Performance remains similar across both allocation schemes (constant and decreasing batch sizes), with decreasing batch sizes showing slight advantages at very small sample sizes ($N_S = 3k$), indicating that allocation strategy has minimal impact on our approach's effectiveness.

Notably, performance remains similar across both allocation schemes (constant and decreasing batch sizes), demonstrating robustness to initial sample size choices. The decreasing batch size configuration — which begins with larger initial batches — shows only marginal advantages at very small total sample sizes ($N_S = 3k$). This indicates that even modest initial samples (e.g., 300 customers in the constant allocation scheme) suffice for effective EPL estimation when using our multi-stage sequential approach, as the iterative refinement of EPL estimates compensates for initial uncertainty. Thus, firms need not be overly concerned with determining an optimal initial sample size; our method maintains strong performance across a range of initialization strategies.

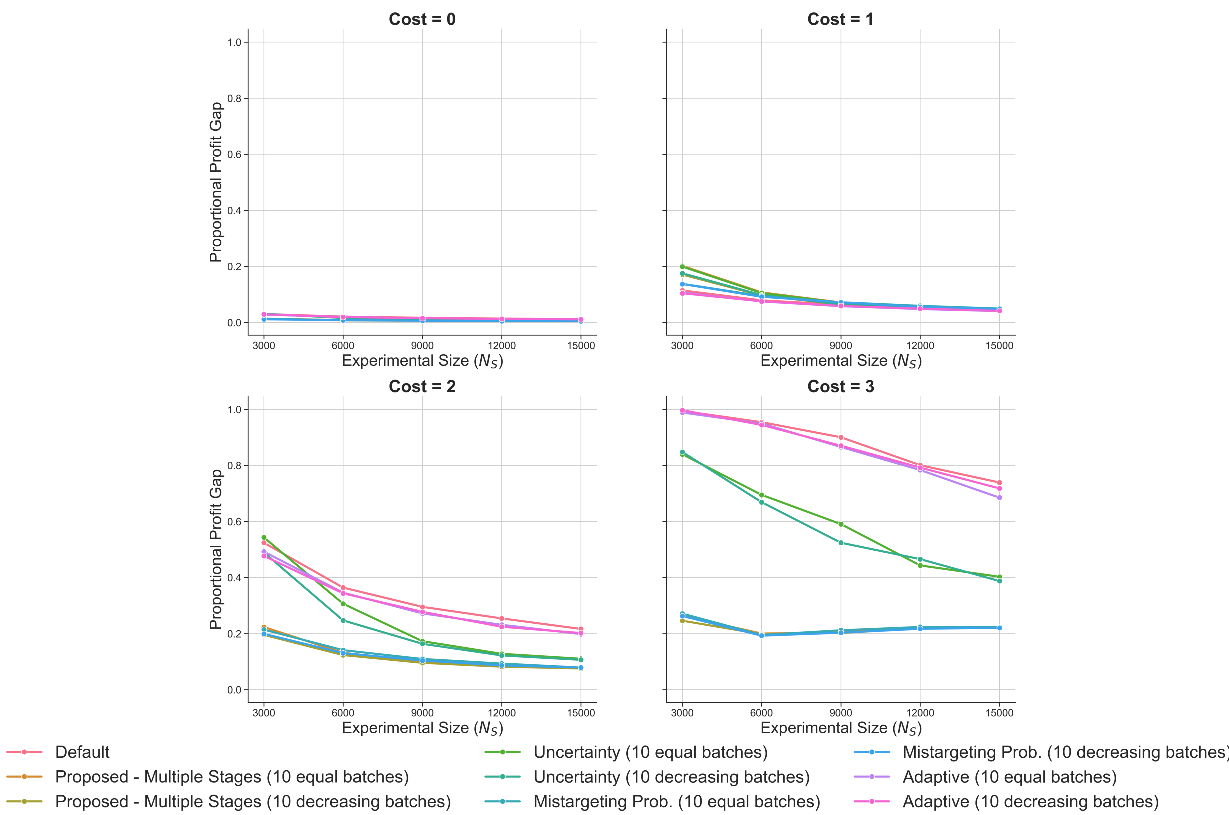


Figure W-3: Proportional Profit Gaps under Homoskedastic Errors with Alternative Sample Allocation Schemes

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

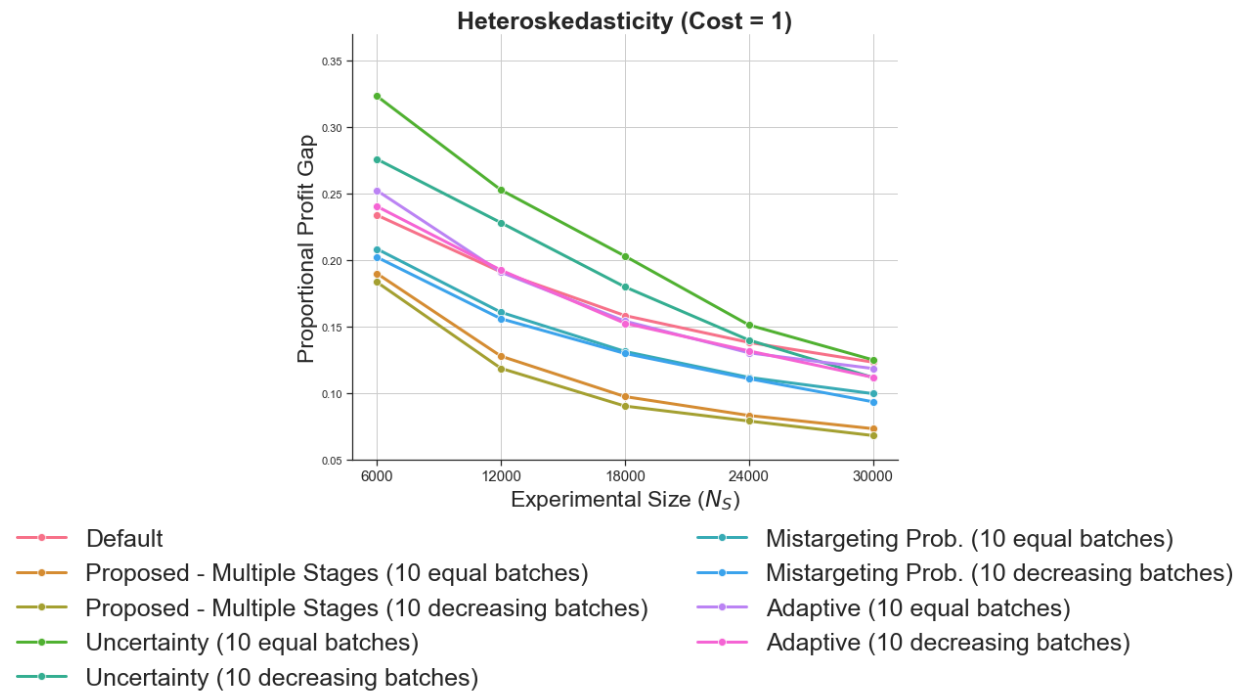


Figure W-4: Proportional Profit Gaps under Heteroskedastic Errors with Alternative Sample Allocation Schemes

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

D.3.2. 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: $r \in \{0.5, 0.7, 0.9\}$.

Figures W-5 and W-6 show the proportional profit gaps of targeting policies learned by multi-stage and two-stage designs under different first-stage sampling proportions ($r \in \{0.5, 0.7, 0.9\}$) across intervention costs ($c \in \{0, 1, 2, 3\}$ for Scenario 1; $c = 1$ for Scenario 2) and experimental sizes ($N_S \in \{3k, 6k, 9k, 12k, 15k\}$).

Under homoskedastic errors, the two-stage design generally performs comparably to the multi-stage version. When the cost aligns with the mode and sample sizes are small (e.g., $N_S = 3k$), a two-stage design with higher first-stage proportion ($r = 0.9$) performs best, as more accurate initial EPL estimates are crucial for effectively identifying consequential customers when they cluster near the threshold. Conversely, when the cost deviates from the mode, a two-stage design with smaller initial samples (e.g., $r = 0.5$) performs better across experimental sizes, as allocating more samples to the second stage enables more intensive recruitment of the scarce consequential customers identified in the first stage.

Under heteroskedastic errors, the ability to recruit consequential customers in the second stage becomes more critical. Two-stage designs with smaller initial samples ($r \in \{0.5, 0.7\}$) underperform the multi-stage approach at small sample sizes (e.g., $N_S \leq 6k$) but achieve comparable performance once sample sizes reach $9k$ or larger, as sufficient data accumulates to identify high-impact, high-uncertainty customers. In contrast, the two-stage design with $r = 0.9$ consistently underperforms across all sample sizes, as the limited second-stage sample restricts intensive sampling of the consequential customers identified in the first stage. These patterns underscore that when prediction difficulty correlates with economic impact, preserving adequate second-stage capacity to concentrate sampling on consequential customers matters more than maximizing initial estimation accuracy.

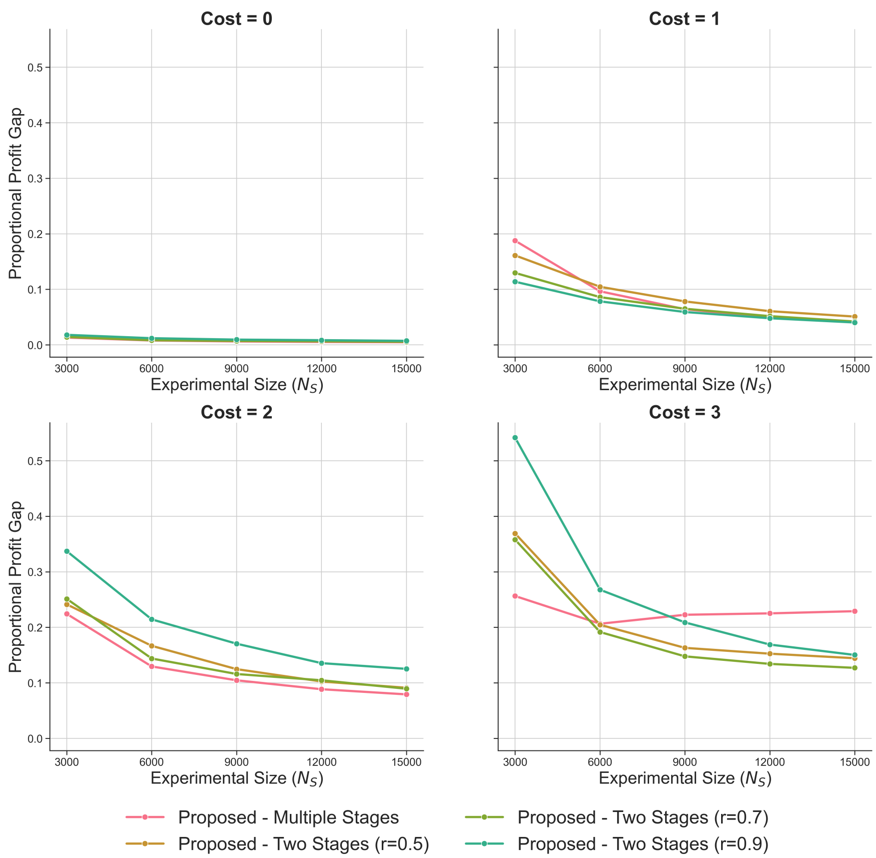


Figure W-5: Proportional Profit Gaps of the Two-Stage Design under Homoskedastic Errors with Alternative Configurations

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

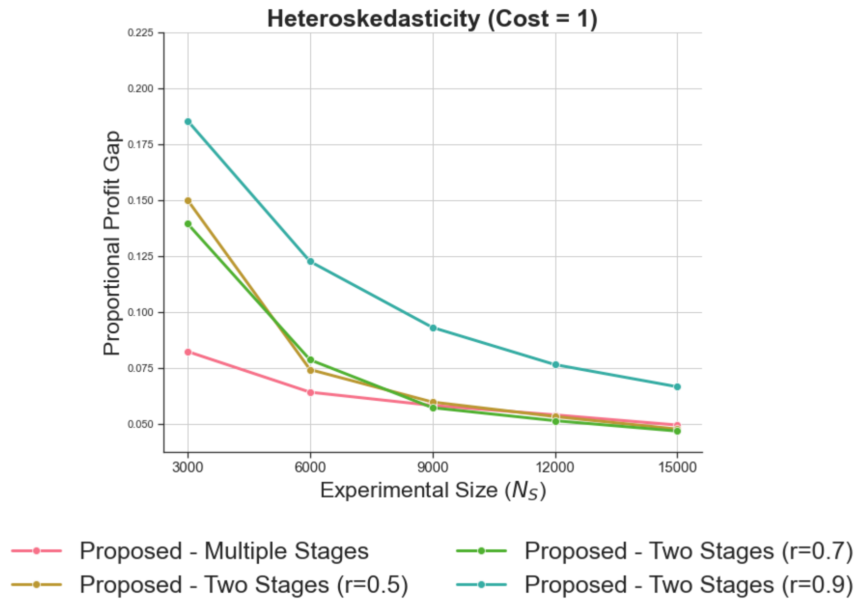


Figure W-6: Proportional Profit Gaps of the Two-Stage Design under Heteroskedastic Errors with Alternative Configurations

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

D.4. Additional Results for Strategic Treatment Assignment

In this appendix, we present additional results examining whether combining our customer selection strategy with adaptive treatment assignment can further enhance performance. Our baseline approach selects customers based on EPL scores but assigns them to treatment and control randomly with equal probability. A natural question is whether strategically allocating customers to treatment arms based on uncertainty could improve CATE estimation accuracy and subsequently yield better targeting policies.

To address this question, we implement a hybrid approach that combines our EPL-based customer selection with adaptive treatment assignment. Specifically, following Kato et al. (2024), we assign selected customers to treatment arms with probabilities proportional to the uncertainty in their potential outcomes:

$$P_{S^{b-1}}(W_i = w|X_i) = \frac{\hat{\sigma}^w_{S^{b-1}}(X_i)}{\hat{\sigma}^0_{S^{b-1}}(X_i) + \hat{\sigma}^1_{S^{b-1}}(X_i)}, \quad w \in \{0, 1\} \tag{W-5}$$

where $\hat{\sigma}^w_{S^{b-1}}(X_i)$ denotes the estimated standard deviation of potential outcome $Y_i(w)$ based on data collected in the first $b-1$ batches. This assignment strategy concentrates observations in the treatment arm where each customer's outcome is most uncertain, potentially improving CATE estimation efficiency. We evaluate whether this combined approach outperforms our baseline method that uses EPL-based selection with random treatment assignment.

Figures W-7 and W-8 show the proportional profit gaps of targeting policies learned by our baseline approach (EPL selection with random assignment) versus the hybrid approach (EPL selection with adaptive assignment) across intervention costs ($c \in \{0, 1, 2, 3\}$ for Scenario 1; $c = 1$ for Scenario 2) and experimental sizes ($N_S \in \{3k, 6k, 9k, 12k, 15k\}$). Across both scenarios and all sample sizes, the hybrid approach with adaptive treatment assignment provides negligible improvements over our baseline approach with random assignment. These findings suggest that strategic customer selection is the primary driver of performance gains in our method, while the marginal benefit of adaptive treatment assignment is minimal.

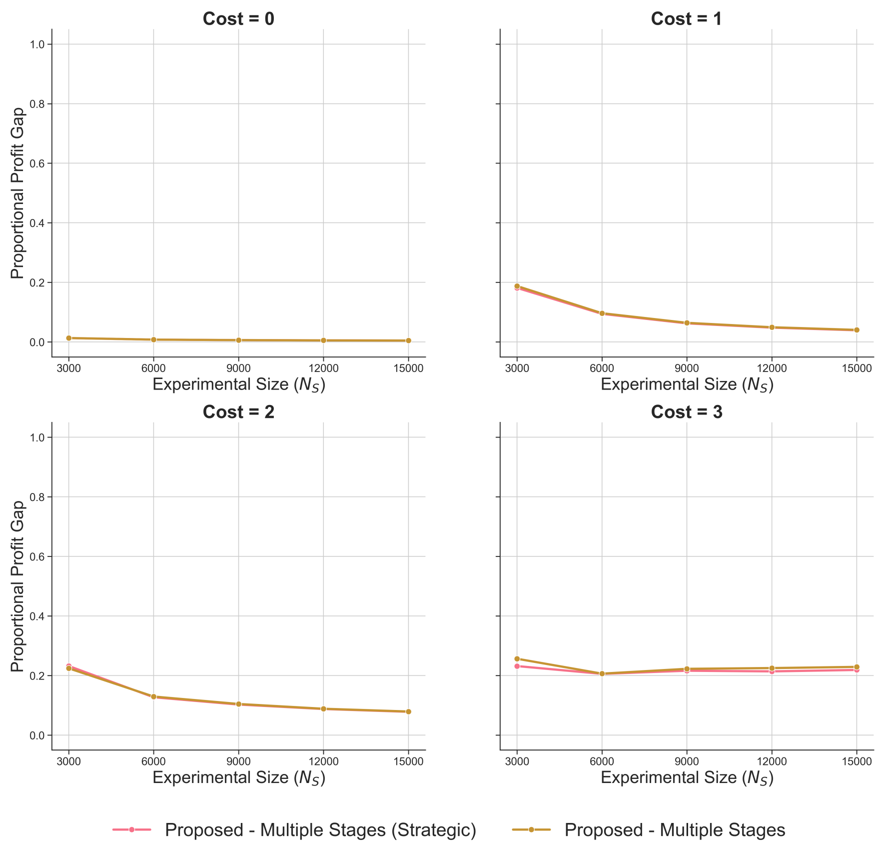


Figure W-7: Proportional Profit Gaps with Strategic vs. Random Treatment Assignment under Homoskedastic Errors
We report the average value of the proportional profit gap across 100 replications. Each line corresponds to an experimental approach.

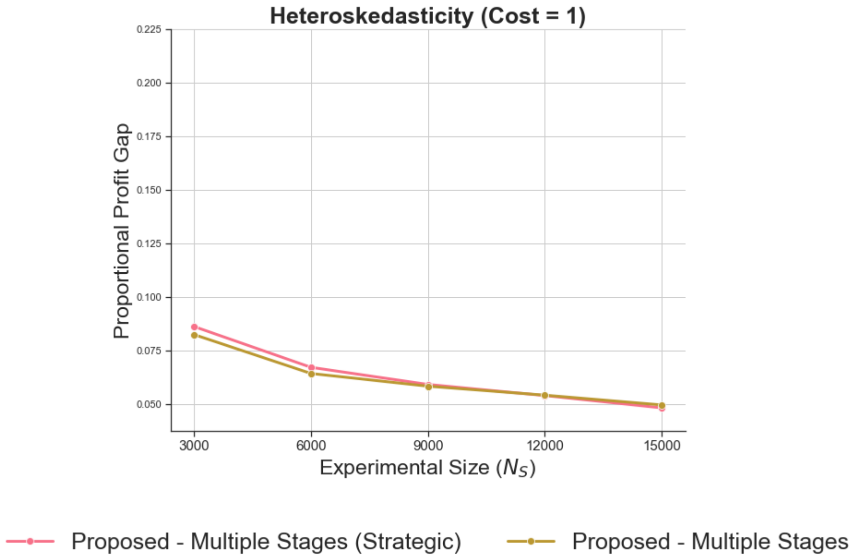


Figure W-8: Proportional Profit Gaps with Strategic vs. Random Treatment Assignment under Heteroskedastic Errors

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

D.5. Additional Results for Stochastic Sampling

In this appendix, we present additional results examining the robustness of our deterministic sampling approach to potential concentration in suboptimal regions of the covariate space. A natural concern with greedy selection of highest-EPL customers is whether insufficient initial exploration might trap the algorithm in local regions, preventing discovery of consequential customers in underrepresented segments. To address this concern, we implement stochastic variants of our method that introduce explicit randomization into customer selection while still prioritizing high-EPL customers. Specifically, we use the Gumbel-top-k method (Kool et al. 2019), which adds random noise to EPL scores before selection. We evaluate two approaches: (1) a *fully stochastic* variant that applies randomized selection throughout the entire experiment, and (2) *hybrid variants* that use stochastic sampling only for the first 600, 1,200, or 1,800 customers to enhance early exploration, then switch to deterministic selection for remaining batches. This analysis assesses whether adding stochasticity, either throughout or concentrated in early stages, improves performance by enhancing coverage of the covariate space, or whether our deterministic method's uncertainty quantification already provides sufficient exploration in the covariate space.[3]

[3] We implement the Gumbel-top-k method by adding Gumbel noise to EPL scores and selecting the top-k customers with highest perturbed scores in each batch.

Figures W-9 and W-10 show the proportional profit gaps of targeting policies learned by deterministic and stochastic sampling approaches across intervention costs ($c \in \{0, 1, 2, 3\}$ for Scenario 1; $c = 1$ for Scenario 2) and experimental sizes ($N_S \in \{3k, 6k, 9k, 12k, 15k\}$). Across both scenarios and all sample sizes, early-stage stochastic variants (600, 1,200, and 1,800 customers) provide negligible improvements over the deterministic approach, while the fully stochastic variant consistently underperforms. These findings confirm that our deterministic approach already provides sufficient coverage of the covariate space through its uncertainty-aware design: underrepresented regions naturally receive higher EPL scores due to elevated estimation uncertainty, eliminating the need for explicit randomization to avoid concentration in suboptimal regions.

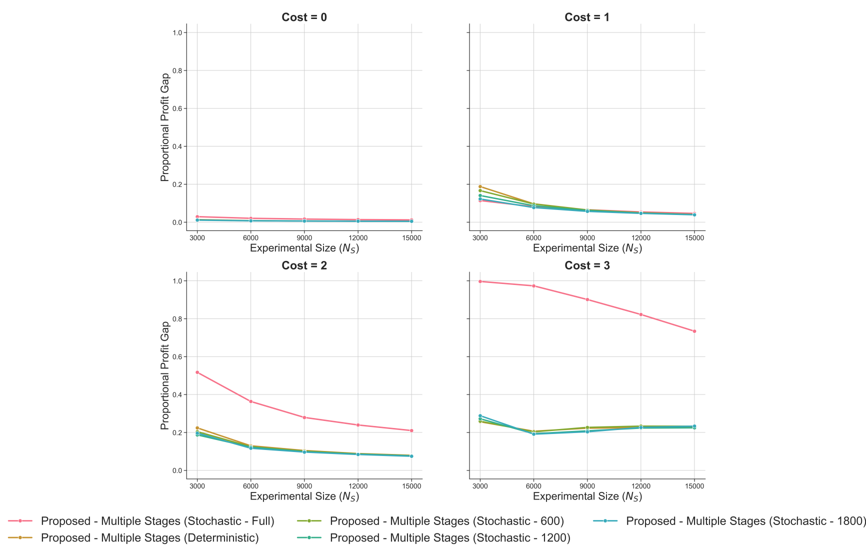


Figure W-9: Proportional Profit Gaps with Stochastic vs. Deterministic Sampling under Homoskedastic Errors with Stochastic Sampling

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

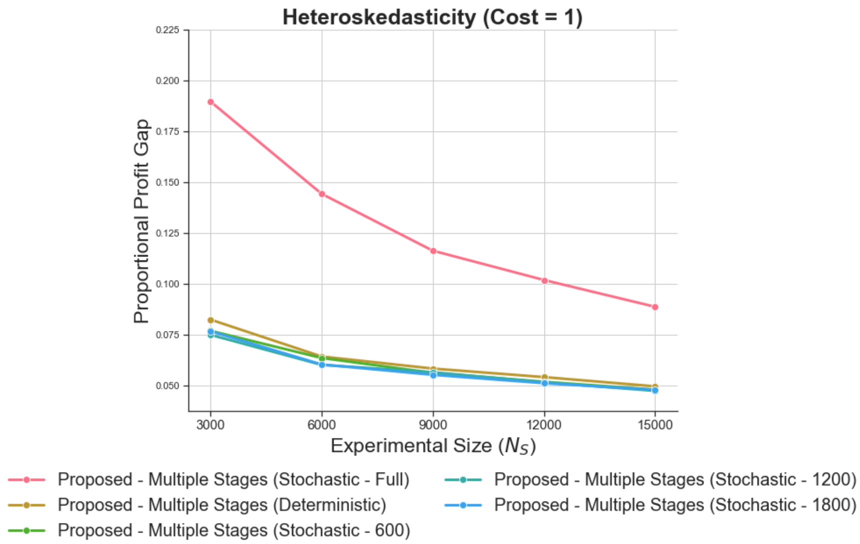


Figure W-10: Proportional Profit Gaps with Stochastic vs. Deterministic Sampling under Heteroskedastic Errors
We report the average value of the proportional profit gap across 100 replications. Each line corresponds to an experimental approach.

D.6. 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 ($r \in \{0.5, 0.7, 0.9\}$).

D.6.1. 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-11 visualizes the relationship between the CATE distribution and the intervention costs.

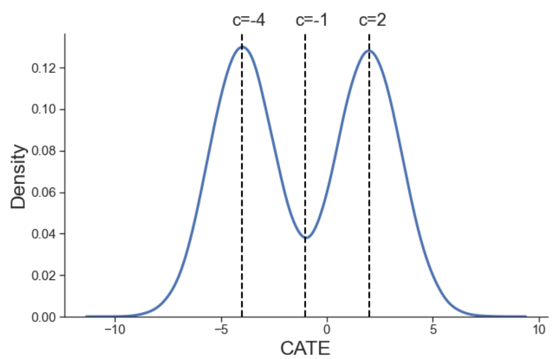


Figure W-11: CATE Distribution and Intervention Costs: Bimodal Distribution with Two Equal Segments

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-12 and Figure W-13 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 ($N_S \in \{6k, 9k, 12k, 15k\}$) respectively.[4] Interestingly, our approach outperforms all benchmarks across all intervention cost, especially when the sample size is small.[5] This highlights the value and sample efficiency of our approach in cases where consumers can be segmented into multiple groups with similar sizes. [6]

[4]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 $N_S = 3k$ for this case.

[5]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.

[6]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

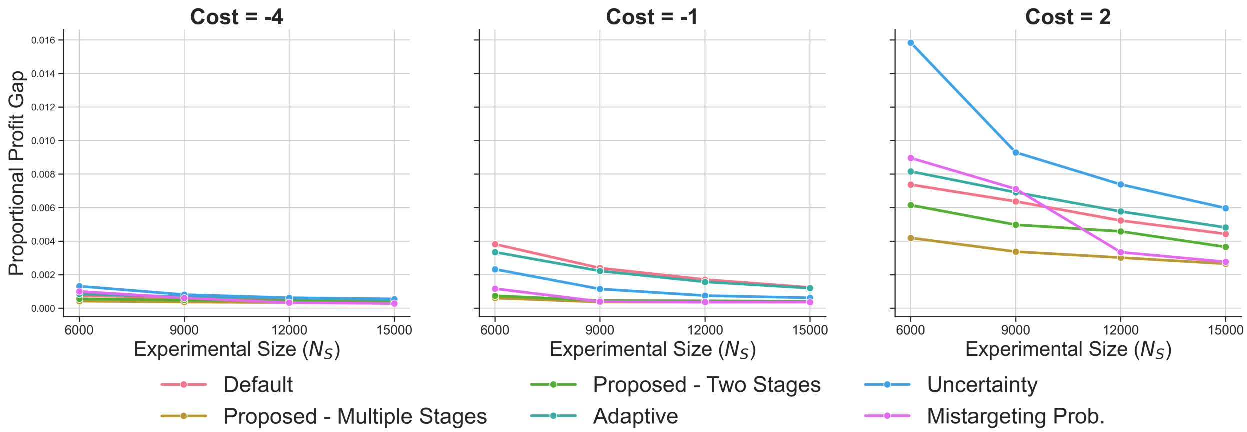


Figure W-12: Proportional Profit Gaps: Bimodal Distribution with Two Equal Segments
We report the average value of the proportional profit gap across 100 replications. Each line corresponds to an experimental approach.

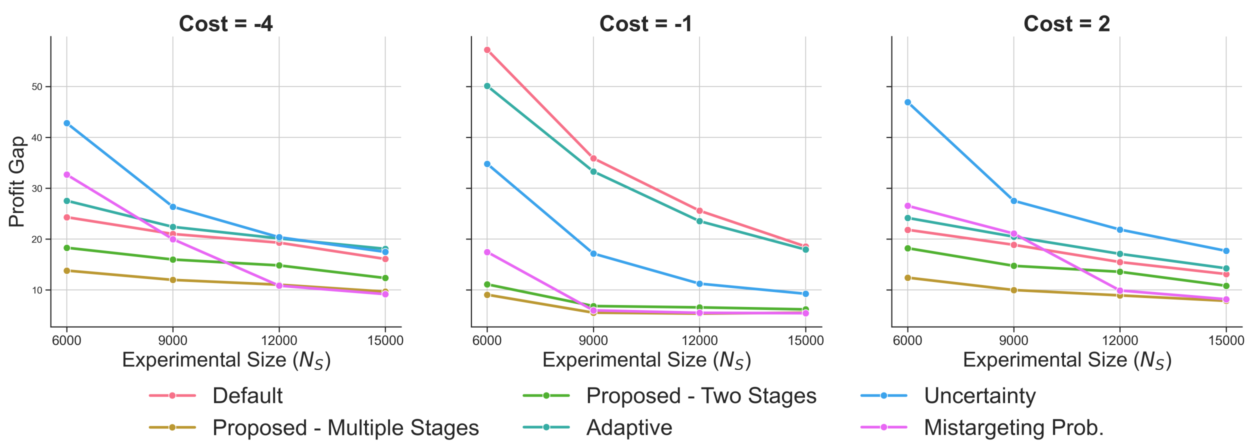


Figure W-13: Profit Gaps: Bimodal Distribution with Two Equal Segments
We report the average value of the proportional profit gap across 100 replications. Each line corresponds to an experimental approach.

Figure W-14 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.

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.

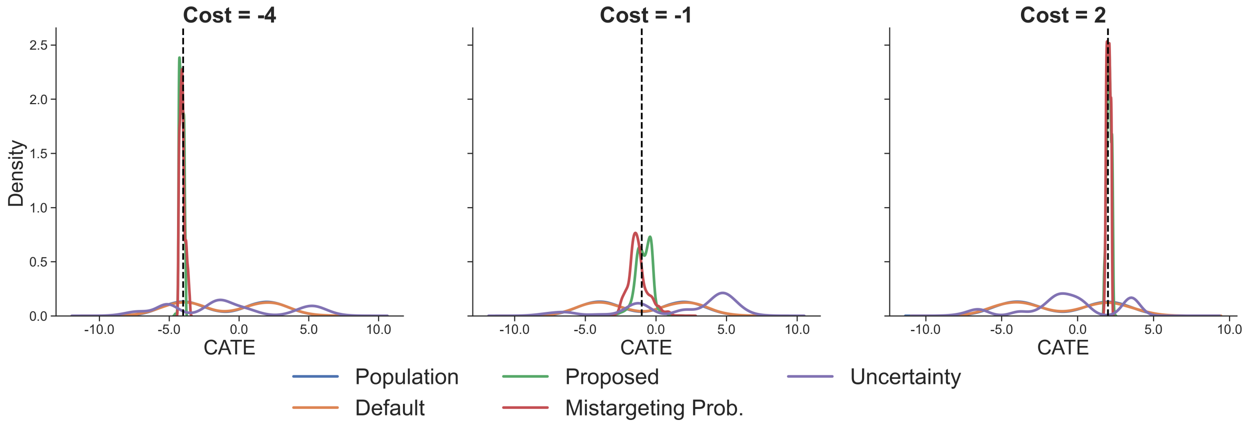


Figure W-14: CATE Distributions of Customers Sampled by Different Approaches: Bimodal Distribution with Two Equal Segments
Each line corresponds to the CATE distribution of the customers sampled by different approaches. The dashed line represents the intervention cost, which is also the decision threshold.

D.6.2. 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-15 illustrates the relationship between the CATE distribution and the intervention costs.

Figure W-16 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 ($N_S \in \{6k, 9k, 12k, 15k\}$).[7]

[7] 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 $N_S = 3k$ for this case.

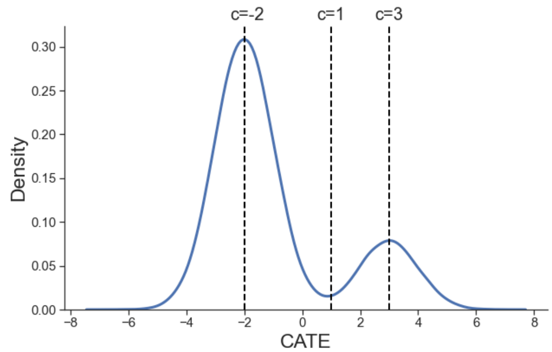


Figure W-15: CATE Distribution and Intervention Costs: Bimodal Distribution with Two Unequal Segments
Each dashed line corresponds to a different intervention cost c. $c = -2$ aligns with the larger peak and $c = 3$ aligns with the smaller peak of the CATE distribution. $c = 1$ represents the valley of the CATE distribution.

As shown in the graph, our approach performs comparably to the `Mistargeting Prob.` approach and generally surpasses the three benchmarks when the cost is at the valley and the smaller peak of the CATE distribution. This finding is qualitatively similar to the ones in Section 5.4.[8]

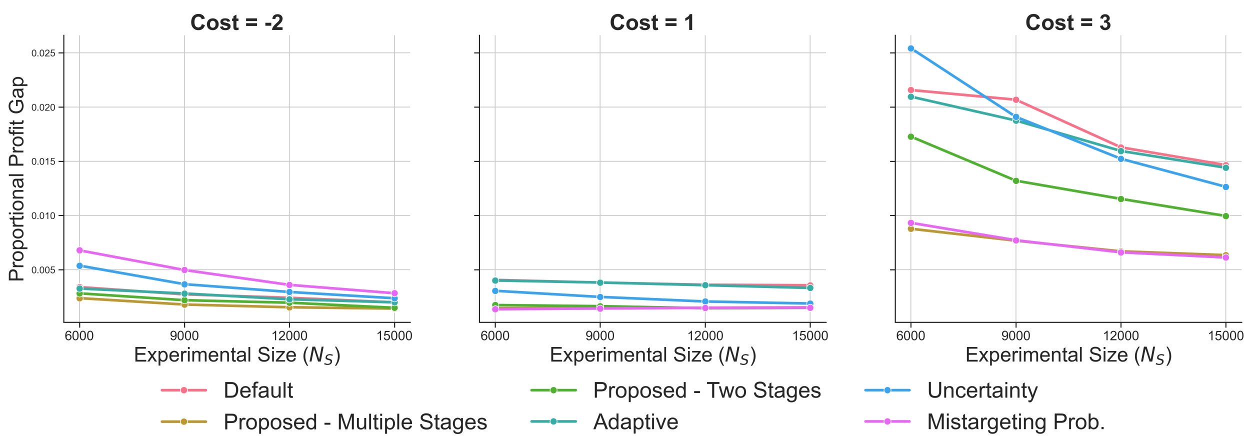


Figure W-16: Proportional Profit Gaps of Different Experimental Designs: Bimodal Distribution with Two Unequal Segments
We report the average value of the proportional profit gap across 100 replications. Each line corresponds to an experimental approach.

Figure W-17 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.

[8] 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.

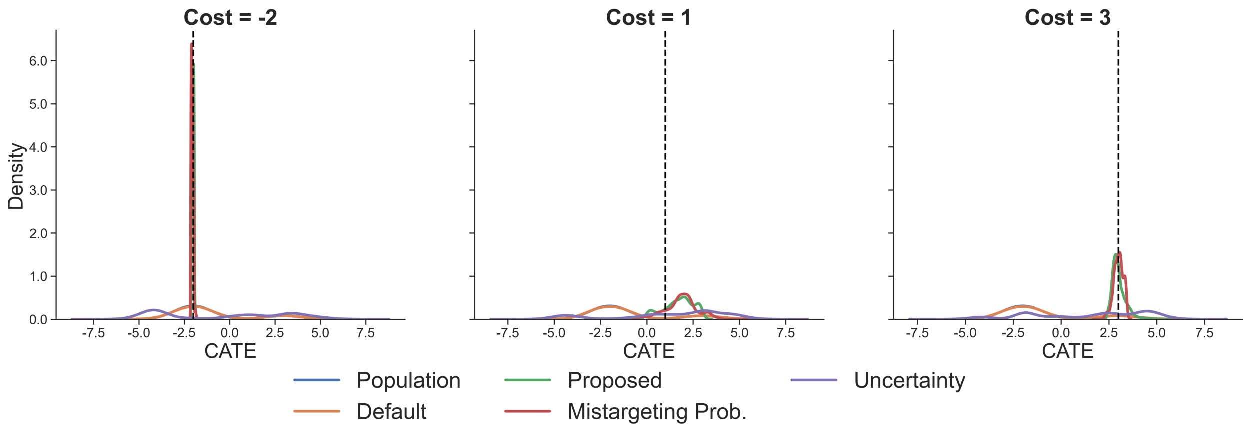


Figure W-17: CATE Distributions of Customers Sampled by Different Approaches: Bimodal Distribution with Two Unequal Segments
Each line corresponds to the CATE distribution of the customers sampled by different approaches. The dashed line represents the intervention cost, which is also the decision threshold.

E. Further Details of Empirical Application

E.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.[9] The results indicate proper randomization, with no significant differences observed between the treatment and control groups across most variables and weeks.

E.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.

[9]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

E.3. Implementation Details

E.3.1. 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.

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	–	–

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, chosen based on cross-validation.

E.3.2. 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, chosen based on cross-validation.

E.3.3. 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, chosen based on cross-validation.

E.3.4. Full (A/B Test with Entire Customer Base)

For each customer in the customer base $\mathcal{I}$, we use the treatment assignment from the original data as the final treatment assignment in our experimentation. or 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, chosen based on cross-validation.

E.3.5. 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 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, chosen based on cross-validation. The uncertainty (standard error) of the CATE estimates $\sigma(\hat{\tau}_{S^{b-1}}(x))$ is estimated using the bootstrap-of-little-bag approach (Athey et al. 2019) implemented in the package.

E.3.6. Mistargeting Probability

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 mistargeting probability 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, chosen based on cross-validation. The mistargeting probability is estimated by the formula $\Pr\left(\hat{\pi} S^{b-1}(x) \neq \pi^*(x)\right) = \Phi\left(\frac{-|\hat{\tau}_{S^{b-1}}(x) - c|}{\hat{\sigma}_{S^{b-1}}(x)}\right)$ where $\hat{\sigma}_{S^{b-1}}(x)$ is estimated using the bootstrap-of-little-bag approach (Athey et al. 2019) implemented in the package.

E.4. Additional Results for Direct Policy Learning

In this appendix, we present additional results examining the performance of our approach when the final targeting policy is estimated using a direct policy learning method. Specifically, we implement Policy Forest (Athey and Wager 2021), which directly learns the optimal policy by maximizing expected outcomes without explicitly estimating conditional average treatment effects. This analysis assesses whether our method's advantages persist when paired with policy learning methods that optimize decision rules directly rather than through predicted treatment effects, and thus being generalizable beyond Causal Forest.[10]

Figure W-18 and Figure W-19 present the results of targeting policies learned from different sampling strategies with direct policy learning methods. The results are qualitatively similar to those in Section 6. In the telecom application, our approach ourperforms `Default` and `Mistargeting Prob.` and performs comparably to `Uncertainty`. In contrast, in the Starbucks application, our approach performs similarly to `Mistargeting Prob.` and outperforms other benchmarks in 50% discount scenario. These findings confirm that EPL-based sampling effectively improves targeting performance across different policy learning approaches by strategically prioritizing consequential customers. Interestingly, consistent with Huang and Ascarza (2024), we also find that causal forest can actually generate more profitable targeting policies than direct policy learning, indicat-

[10]We implement the Policy Forest method using the `econML` package in Python. The model comprises 300 trees with maximum depths not exceeding 5, chosen based on cross-validation.

ing that policy learning doesn't always benefit from its ability to directly optimize the objective function.

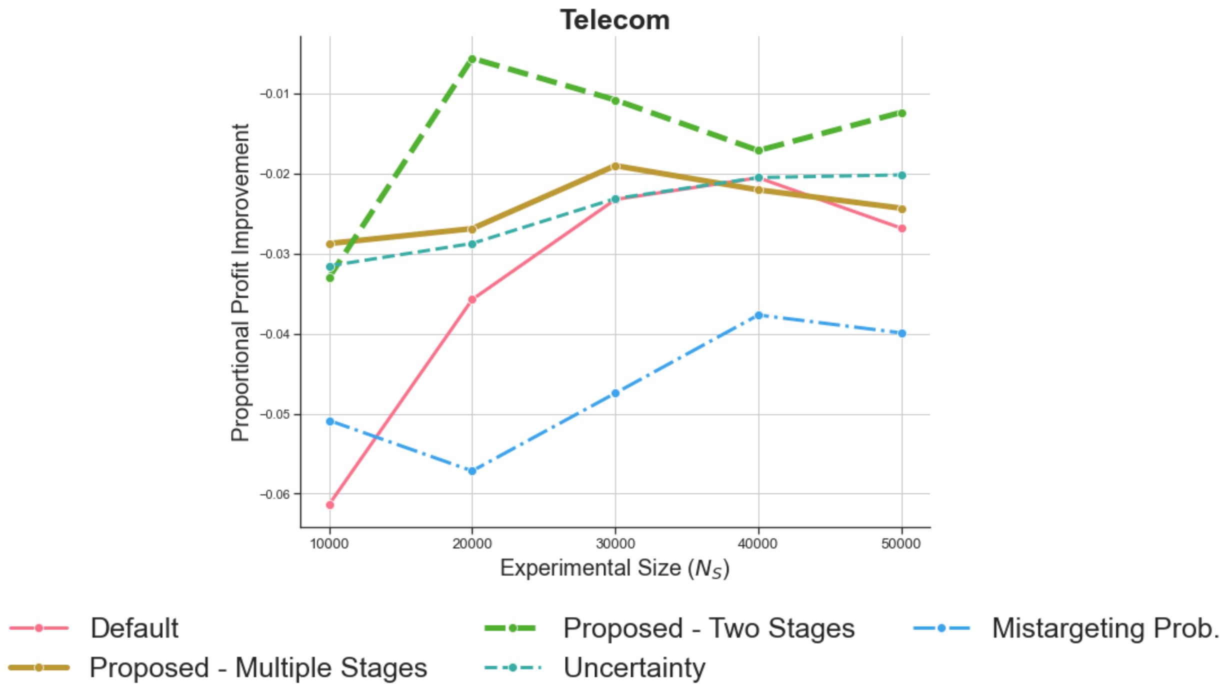


Figure W-18: Performance of Targeting Policies with Direct Policy Learning
We report the average value of the proportional profit improvement relative to the uniform policy across 100 replications. Each line corresponds to an experimental approach.

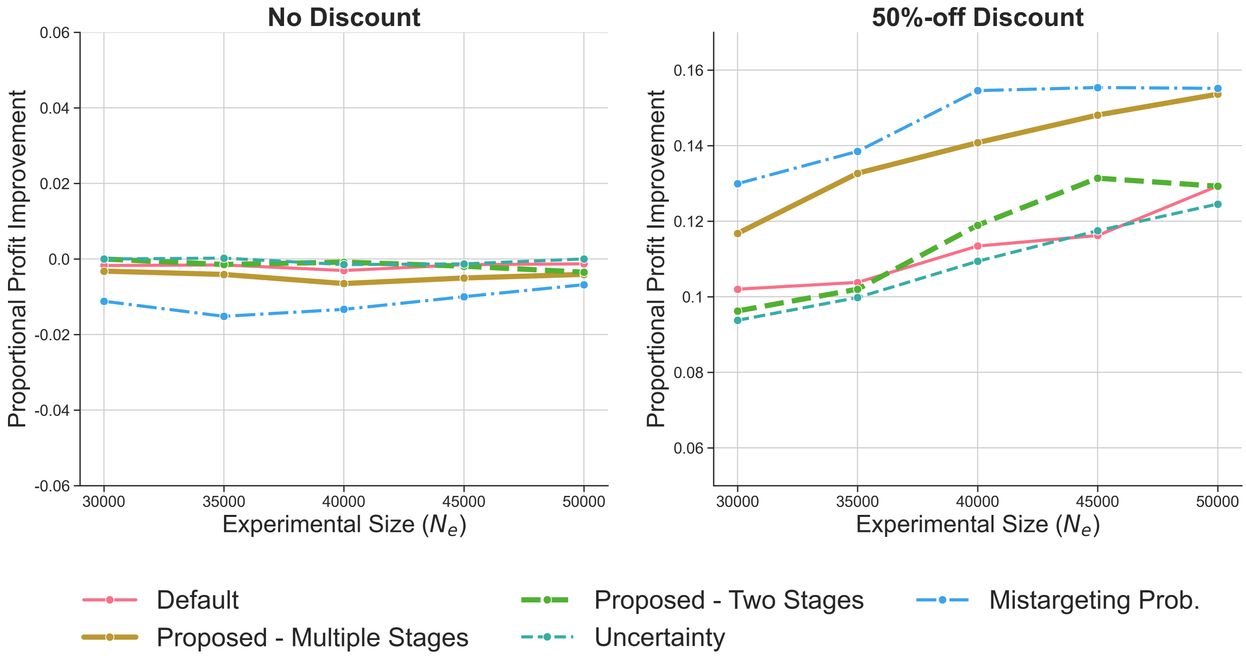


Figure W-19: Performance of Targeting Policies with Direct Policy Learning
We report the average value of the proportional profit improvement relative to the uniform policy across 100 replications. Each line corresponds to an experimental approach.

E.5. 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.

E.5.1. Alternative Sample Allocation Schemes for Multi-Stage Design

Figure W-20 and Figure W-21 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. 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.

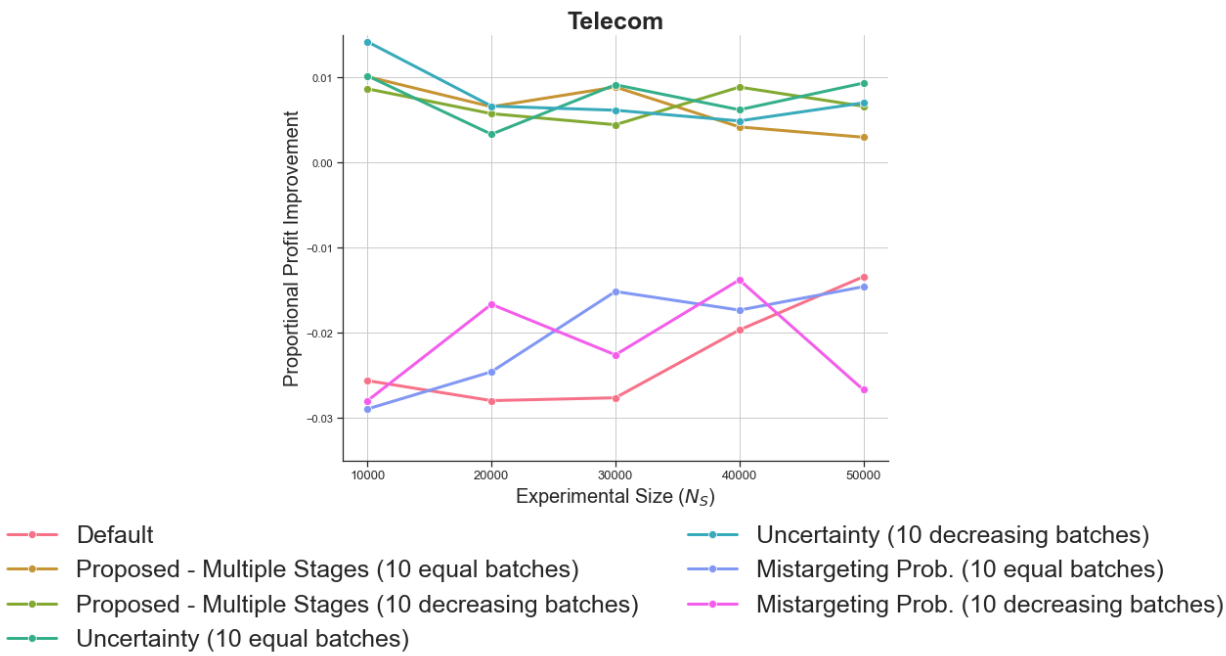


Figure W-20: Performance of Targeting Policies Learned from Different Multi-Stage Designs (Telecommunication)

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.

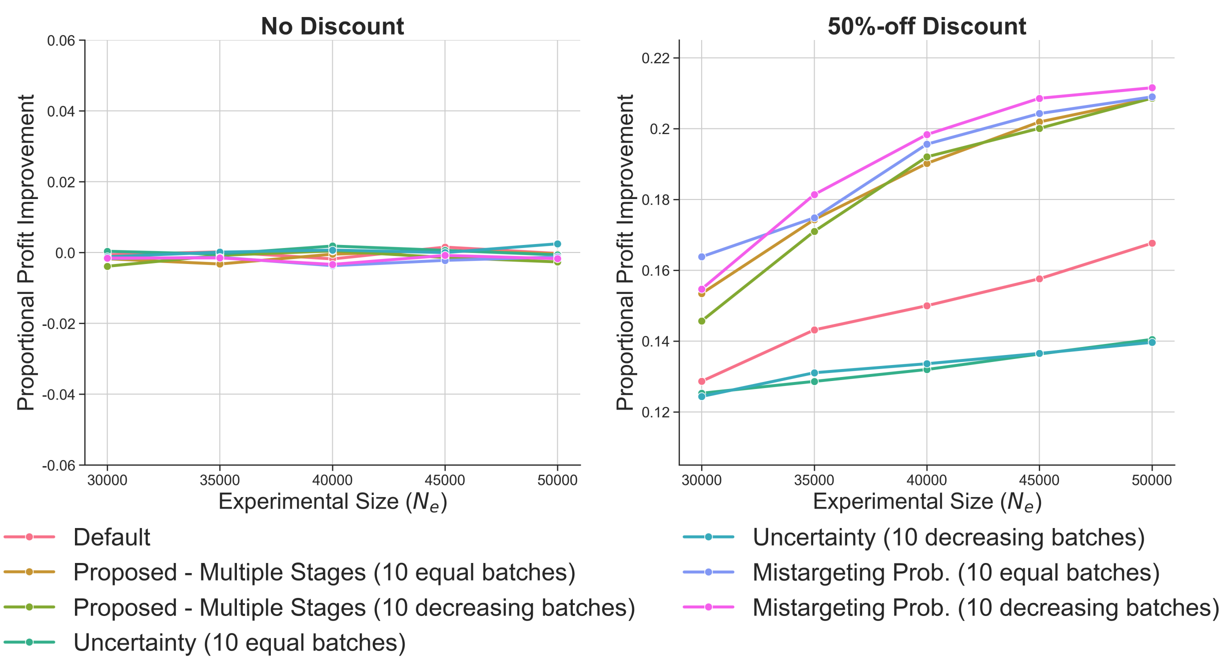


Figure W-21: Performance of Targeting Policies Learned from Different Multi-Stage Designs (Starbucks)

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.

E.5.2. Alternative Configurations for Two-Stage Design

Figure W-22 and Figure W-23 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.

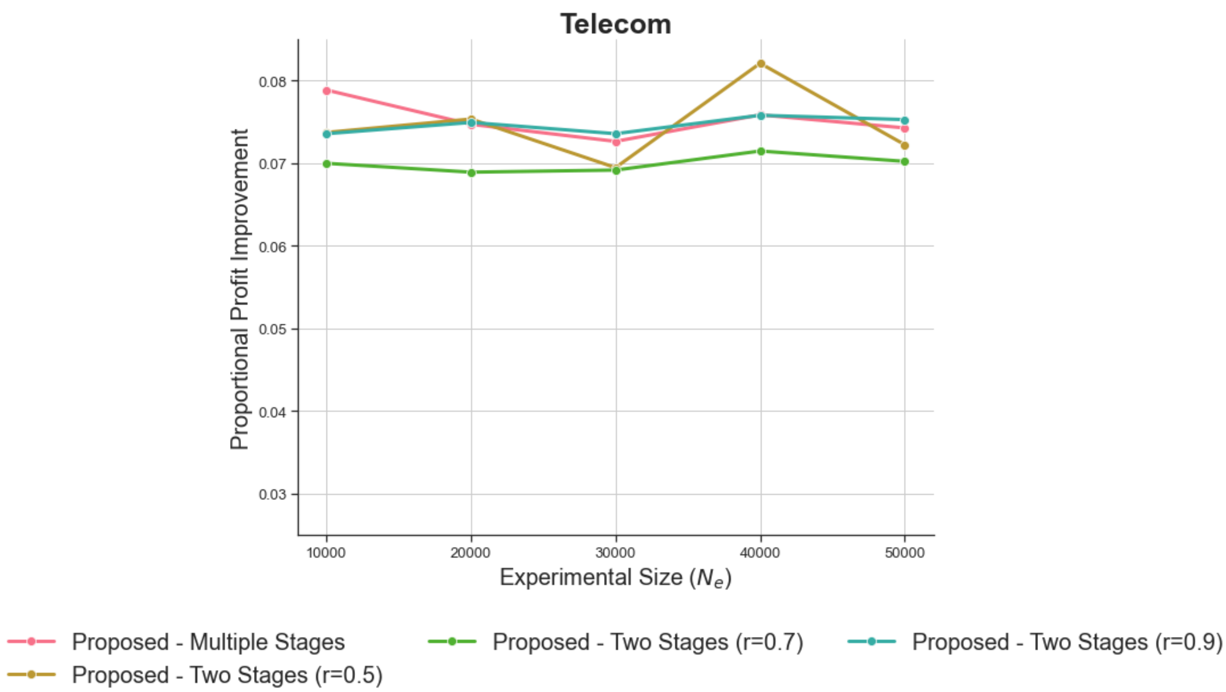


Figure W-22: Performance of Targeting Policies Learned from Different Two-Stage Designs (Telecommunication)
We report the average value of the proportional profit improvement relative to the uniform policy across 100 replications. Each line corresponds to an experimental approach.

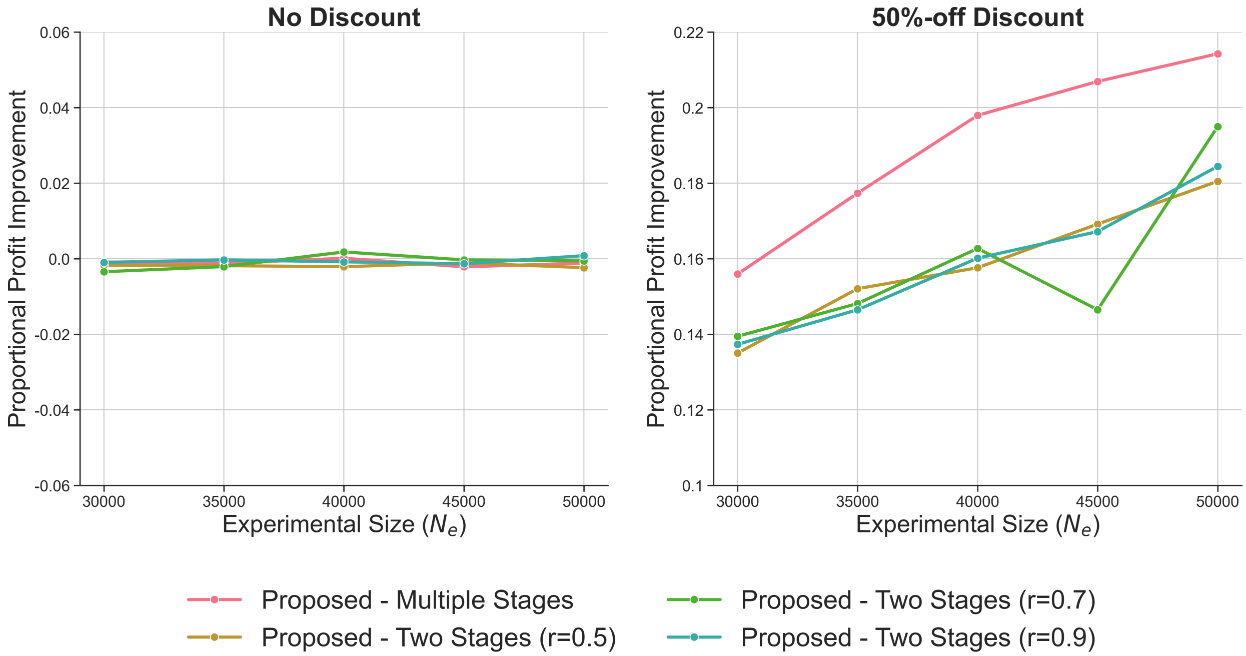


Figure W-23: Performance of Targeting Policies Learned from Different Two-Stage Designs (Starbucks)
We report the average value of the proportional profit improvement relative to the uniform policy across 100 replications. Each line corresponds to an experimental approach.

E.6. Additional Results for Stochastic Sampling

In this appendix, we present additional results examining the robustness of our deterministic sampling approach to potential concentration in suboptimal regions of the covariate space. Specifically, we use the Gumbel-top-k method (Kool et al. 2019), which adds random noise to EPL scores before selection. We evaluate two approaches: (1) a *fully stochastic* variant that applies randomized selection throughout the entire experiment, and (2) *hybrid variants* that use stochastic sampling only for the first $1,000$, $2,000$, or $5,000$ customers to enhance early exploration, then switch to deterministic selection for remaining batches. This analysis assesses whether adding stochasticity, either throughout or concentrated in early stages, improves performance by enhancing coverage of the covariate space, or whether our deterministic method's uncertainty quantification already provides sufficient exploration in the covariate space.[11]

Figures W-24 and W-25 show performance of targeting policies learned by deterministic and stochastic sampling approaches for both applications. Across both applications and all sample sizes, early-stage stochastic variants ($1,000$, $2,000$, and $5,000$ customers) provide negligible im-

[11]We implement the Gumbel-top-k method by adding Gumbel noise to EPL scores and selecting the top-k customers with highest perturbed scores in each batch.

provements over the deterministic approach, while the fully stochastic variant underperforms in Starbucks 50% discount scenario. These findings confirm that our deterministic approach already provides sufficient coverage of the covariate space through its uncertainty-aware design: under-represented regions naturally receive higher EPL scores due to elevated estimation uncertainty, eliminating the need for explicit randomization to avoid concentration in suboptimal regions.

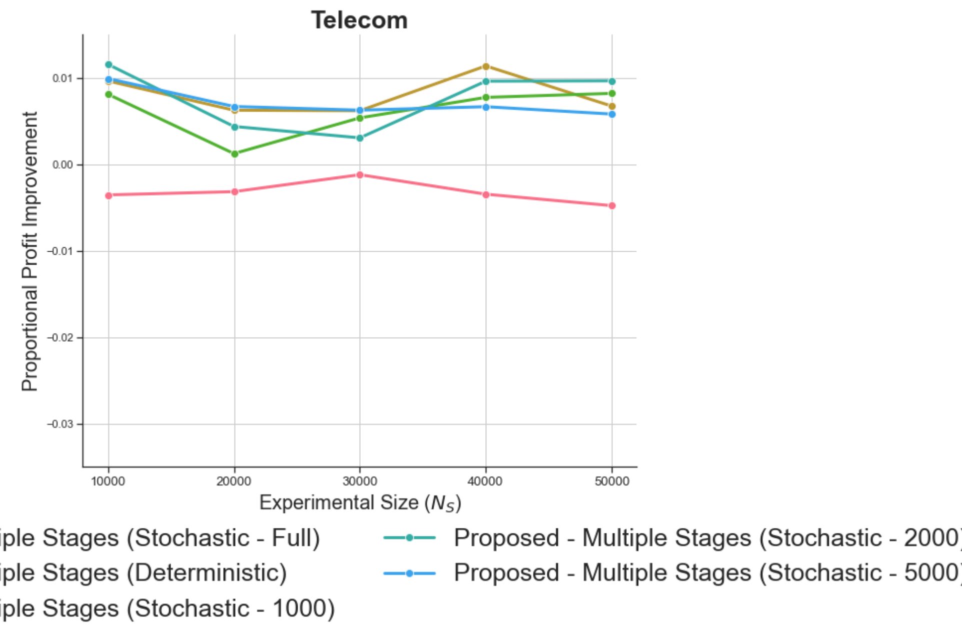


Figure W-24: Performance of Targeting Policies with Stochastic vs. Deterministic Sampling
We report the average value of the proportional profit improvement relative to the uniform policy across 100 replications. Each line corresponds to an experimental approach.

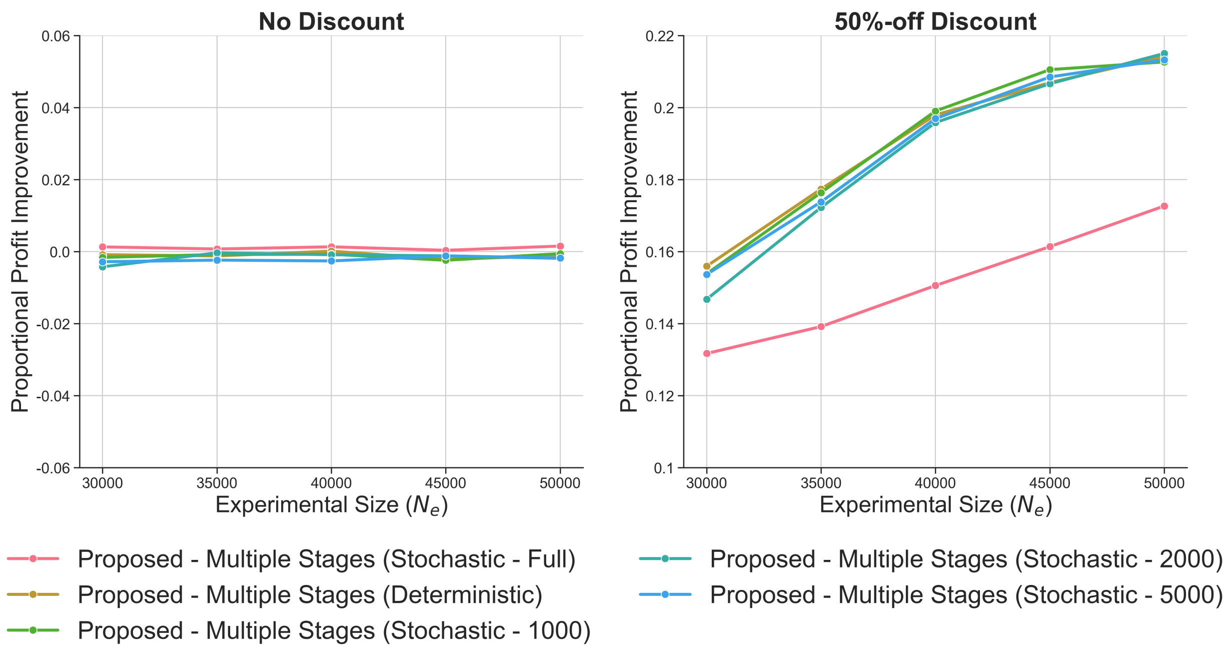


Figure W-25: Performance of Targeting Policies with Stochastic vs. Deterministic Sampling
We report the average value of the proportional profit improvement relative to the uniform policy across 100 replications. Each line corresponds to an experimental approach.

F. Discussion of Potential Extensions

F.1. Early Stopping Rule

While our framework assumes a pre-specified maximum sample size N_S, firms can adopt an *early stopping rule* to terminate the experiment earlier when additional data is unlikely to meaningfully improve the targeting policy. We propose two complementary criteria for determining when to stop.

Absolute threshold. The firm can terminate the experiment when the maximum EPL score across remaining customers falls below a pre-specified threshold $\eta > 0$:

$$\max_{i \notin S^b} \hat{\ell}_{S^b}(\mathbf{X}_i) < \eta \tag{W-6}$$

This reflects that no remaining customer would generate meaningful profit loss from mistargeting, making further experimentation unnecessary. The threshold η can be set after the initialization stage by examining the distribution of EPL scores, or based on the firm's tolerance for profit loss.

Convergence criterion. Alternatively, the firm can terminate the experiment when the change in the maximum EPL between successive batches falls below a tolerance level $\epsilon > 0$:

$$\max_{i \notin S^b} \hat{\ell}_{S^b}(\mathbf{X}_i) - \max_{i \notin S^{b-1}} \hat{\ell}_{S^{b-1}}(\mathbf{X}_i) < \epsilon \tag{W-7}$$

This indicates that additional experimental data is unlikely to meaningfully improve targeting profitability, suggesting that the targeting policy has sufficiently stabilized.

The two criteria can be used independently or in combination, with the experiment terminating when either condition is met.

F.2. Extension to Budget Constraint

While our framework focuses on the setting where the firm targets all customers with positive incremental profit (i.e., $d = 0$), in practice firms often face budget constraints that limit the number of customers they can treat. Our approach can naturally accommodate such constraints by introducing an elevated threshold $d > 0$, replacing the decision rule with $\pi^*(\mathbf{X}_i) = \mathbf{1}\{\tau(\mathbf{X}_i) - c(\mathbf{X}_i) > d\}$, where d is chosen to meet the budget. Since EPL prioritizes customers near the decision threshold regardless of where that threshold is set, our sampling framework applies without modification under this extended decision rule.

However, we acknowledge that this is only a partial solution: a fully optimal treatment of budget constraints would require the endogenous determination of d under uncertainty, accounting for the fact that the optimal threshold itself depends on the distribution of treatment effects that has yet to be fully learned. We leave this more complete treatment of budget constraints for future research.

F.3. Extension to Multiple Treatment Arms

While our framework focuses on binary treatment assignments which is consistent with our empirical applications, the underlying principle of EPL-based sampling extends naturally to settings with multiple treatment arms. Specifically, let $\mathcal{W} = \{0, 1, \ldots, K\}$ denote the set of $K + 1$ treatment arms, $\tau_k(\mathbf{X}_i) = \mathbb{E}[Y_i(k) - Y_i(0) \mid \mathbf{X}_i]$ the CATE for arm k relative to control, and $c_k(\mathbf{X}_i)$ the cost of assigning arm k to customer i. Under this set up, the optimal and estimated policies are respectively:

$$\pi^*(\mathbf{X}_i) = \arg\max_{k \in \mathcal{W}} \{\tau_k(\mathbf{X}_i) - c_k(\mathbf{X}_i)\}, \quad \hat{\pi}_S(\mathbf{X}_i) = \arg\max_{k \in \mathcal{W}} \{\hat{\tau}_{k,S}(\mathbf{X}_i) - c_k(\mathbf{X}_i)\} \tag{W-8}$$

where $\hat{\tau}_{k,S}(\mathbf{X}_i)$ denotes the estimated CATE for arm k based on the current experimental sample S. The multi-arm EPL for customer i is then defined as:

$$\ell_S(\mathbf{X}_i) = \mathbb{E}_{\hat{\pi}}\left[\left(\max_{k\in\mathcal{W}}\{\tau_k(\mathbf{X}_i) - c_k(\mathbf{X}_i)\}\right)\cdot \mathbf{1}\{\hat{\pi}_S(\mathbf{X}_i) \neq \pi^*(\mathbf{X}_i)\}\right] \tag{W-9}$$

This measures the expected profit loss from selecting the wrong treatment arm due to estimation uncertainty, and naturally reduces to the binary EPL in Section 4 when $|\mathcal{W}| = 2$. Thus, a greedy algorithm that iteratively selects customers with the highest multi-arm EPL scores prioritizes sampling the customers whose economic consequences are most significant.

F.4. Extension to Real-time Customer Arrival

The proposed framework developed in Section 4 assumes that the firm operates over a predetermined customer base from which it can actively recruit experiment participants. While this assumption is natural in a wide range of marketing applications, it does not cover settings in which customers arrive randomly and sequentially, such as visitors to a website or users of a mobile application. In these settings, the firm has no control over who arrives or when; it can only decide, upon each arrival, whether to include that customer in the experiment.

To address this limitation, we extend our proposed framework to the real-time arrival setting with a simple adaptation. The key insight is that the firm can dynamically apply an EPL threshold to each arriving customer: customers whose EPL exceeds the current threshold are enrolled and randomly assigned to treatment or control, while the rest pass through without being included in the experiment. The threshold is estimated from the customers already enrolled and is updated periodically as the experiment accumulates data.

We acknowledge, however, that this extension is necessarily *partial*. A fully optimal solution would require addressing two substantial methodological challenges beyond our current scope: first, it requires setting dynamic inclusion thresholds under uncertainty about future arrivals; second, finding optimal EPL thresholds would require modeling the arrival process and solving for adaptive cutoffs under budget constraints. We encourage future research to develop more sophisticated solutions tailored to this context. Nevertheless, as we show below, the simple adaptation already outperforms other benchmarks, suggesting that the EPL criterion transfers naturally to real-time settings.

F.4.1. Method: Adaptive EPL Thresholding

We consider a setting where customers arrive one at a time in a random, exogenous order. Upon arrival, the firm observes customer i's covariates $\mathbf{X}_i$ and must immediately decide whether to include them in the experiment. The total experiment budget is N_S.

Before the experiment begins, the firm determines the number of batches B and the size of each batch $\{n_b\}_{b=1}^{B}$ such that $\sum_{b=1}^{B} n_b = N_S$. The first batch S_1 of size n_1 serves as the initialization stage, and each subsequent batch S_b ($b = 2, \ldots, B$) collects n_b newly enrolled customers before the inclusion thresholds are updated.

In the initialization stage, the first n_1 arriving customers form S_1. Each customer in S_1 is randomly assigned to treatment arms $W_i \in \{0, 1\}$ and their outcomes Y_i are observed. A CATE model is then estimated on S_1, and the inclusion threshold η_1 is set at the q-th percentile of $\{\hat{\ell}_{S_1}(\mathbf{X}_i)\}_{i \in S_1}$. Here, $q \in [0, 100]$ controls the selectivity of enrollment: a higher q sets a more stringent threshold, enrolling only customers with the highest EPL scores more slowly, while a lower q enrolls more broadly and fills the experimental budget faster. This is a firm decision that can be updated across batches. As a practical rule of thumb, we recommend starting with a lower q in early batches — when EPL estimates are noisy and broad coverage of the covariate space is valuable — and gradually increasing q in later batches as EPL estimates stabilize. Firms facing time pressure to complete the experiment quickly should favor a uniformly lower q throughout.

For customers arriving in subsequent batches $b = 2, \ldots, B$, the firm computes its EPL score $\hat{\ell}_{S^{b-1}}(\mathbf{X}_i)$ using the CATE model estimated on $S^{b-1} = \cup_{j=1}^{b-1} S_j$, the data collected through all previously completed batches:

$$\text{Enroll customer } i \quad \Longleftrightarrow \quad \hat{\ell}_{S^{b-1}}(\mathbf{X}_i) \geq \eta_{b-1}. \tag{W-10}$$

Similar to the initialization stage, enrolled customers are randomly assigned to treatment arms $W_i \in \{0, 1\}$ and their outcomes Y_i are observed. Once another n_b customers have been enrolled into the experiment, the CATE model is re-estimated on $S^b = \cup_{j=1}^{b} S_j$, with the threshold η_b updated to the q-th percentile of $\{\hat{\ell}_{S^b}(\mathbf{X}_i)\}_{i \in S^b}$. This process repeats until $|S| = N_S$, at which point a final causal forest is estimated on the final experimental data $S = \cup_{b=1}^{B} S_b$ and used to derive the targeting policy.

Pseudocode for the full procedure is provided in Algorithm 3.

Algorithm 3 Adaptive EPL Thresholding for Real-Time Customer Arrival

Input: Experiment budget N_S; number of batches B; batch sizes $\{n_b\}_{b=1}^{B}$; threshold percentile q
Output: Targeting decision $\hat{\pi}_S$
Enroll first n_1 arriving customers as S_1; assign $W_i \in \{0, 1\}$ at random; observe Y_i
Estimate CATE model on S_1; set η_1 at the q-th percentile of $\{\hat{\ell}_{S_1}(\mathbf{X}_i)\}_{i \in S_1}$
for $b = 2, 3, \ldots, B$ **do**
  Observe next arriving customer $i \notin S^{b-1}$ with covariates $\mathbf{X}_i$
  if $\hat{\ell}_{S^{b-1}}(\mathbf{X}_i) \geq \eta_{b-1}$ **then**
    Enroll customer i into S_b; assign $W_i \in \{0, 1\}$ at random; observe Y_i
  end if
  if $|S_b| = n_b$ **then**
    Re-estimate CATE model on $S^b = \cup_{j=1}^{b} S_j$
    Update η_b to the q-th percentile of $\{\hat{\ell}_{S^b}(\mathbf{X}_i)\}_{i \in S^b}$
  end if
end for
Estimate CATE model on $S = \cup_{b=1}^{B} S_b$
Return: $\hat{\pi}_S(x) = \mathbf{1}\{\hat{\tau}_S(x) > c\}$

F.4.2. Simulation Setup

We follow the simulation designs described in Section 5. To operationalize sequential arrival, we allow each customer to arrive sequentially in a random order, and the firm observes them one at a time without control over the arrival sequence. To apply the adapted method described in the previous section, we set the initial batch size to $n_1 = 300$, the threshold percentile to $q = 50$, and the batch size for subsequent updates to $n_b = 300$ for all $b \geq 2$. All methods share the same arrival order and the same initial batch S_1 within each replication, ensuring that performance differences are attributable solely to the inclusion rule. Results are averaged across 20 independent replications.[12]

F.4.3. Simulation Results

Figures W-26 and W-27 show the proportional profit gaps of targeting policies learned by difference sampling approaches under real-time customer arrival across intervention costs ($c \in \{0, 1, 2, 3\}$ for Scenario 1; $c = 1$ for Scenario 2) and experimental sizes ($N_S \in \{3k, 6k, 9k, 12k, 15k\}$). The results are qualitatively similar to those in Section 5.4, with the exception that our approach underperforms `Uncertainty` in the heteroskedastic scenario. These findings confirm that the proposed

[12]We use fewer replications than in the main analysis due to the substantially higher computational cost of the sequential arrival procedure.

EPL criterion transfers naturally to settings where customers arrive sequentially in random order, though further methodological development is needed to improve its performance in this setting.

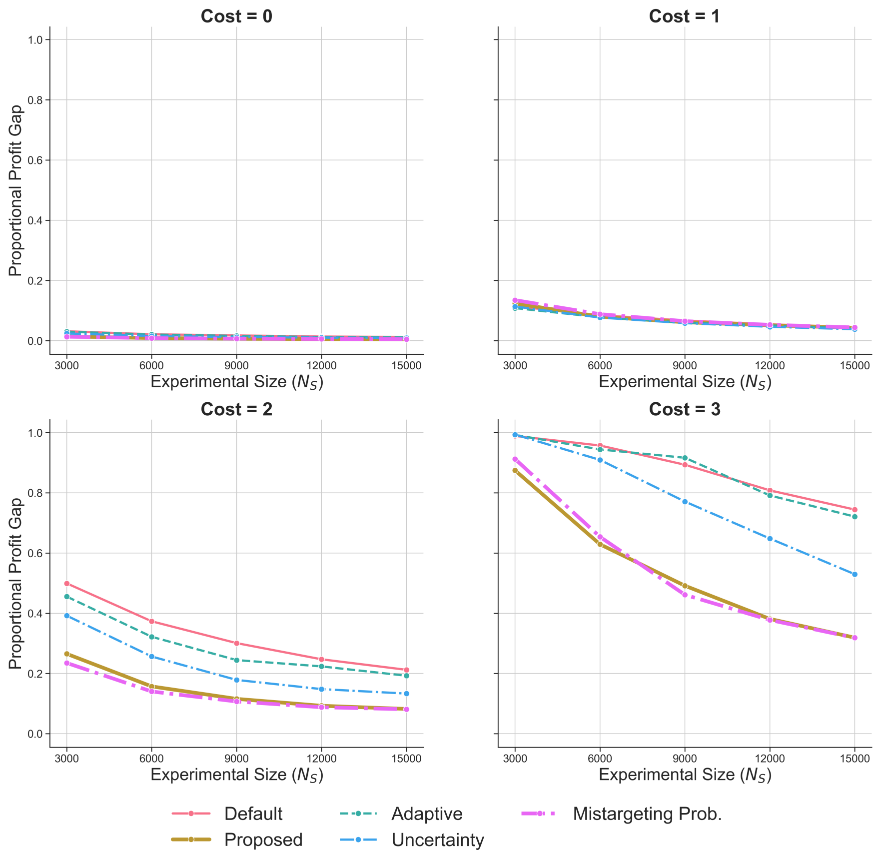


Figure W-26: Proportional Profit Gaps of Real-time Customer Arrival

We report the average value of the proportional profit gap across 20 replications. Each line corresponds to an experimental approach.

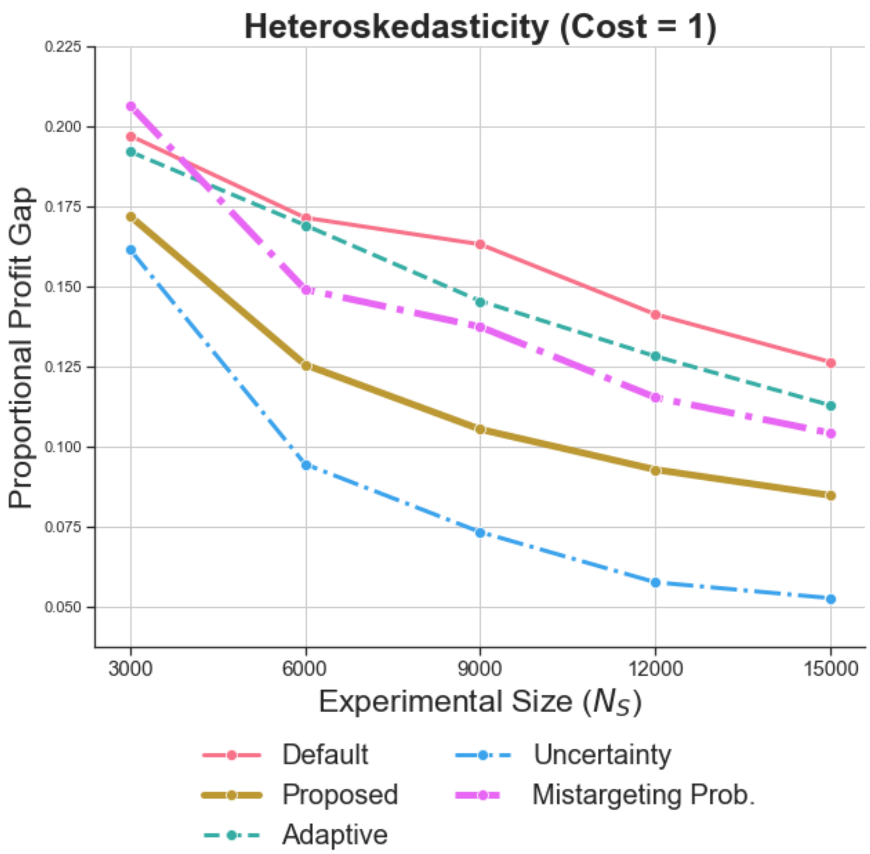


Figure W-27: Proportional Profit Gaps of Real-time Customer Arrival under Heteroskedastic Errors

We report the average value of the proportional profit gap across 20 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.

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].

Elmachtoub, A. N. and Grigas, P. (2022). Smart "Predict, then Optimize". *Management Science*, 68(1):9–26.

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].

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.

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.

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.

Kool, W., van Hoof, H., and Welling, M. (2019). Stochastic Beams and Where to Find Them: The Gumbel-Top-k Trick for Sampling Sequences Without Replacement. Version Number: 2.

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].

Toubia, O. and Hauser, J. R. (2007). **Research Note** —On Managerially Efficient Experimental Designs. *Marketing Science*, 26(6):851–858.

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.

Xie, M. and Singh, K. (2013). Confidence Distribution, the Frequentist Distribution Estimator of a Parameter: A Review. *International Statistical Review*, 81(1):3–39.